PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification [4]: C12N 5/00, 5/02, C12P 19/34 — A1

(11) International Publication Number: WO 88/ 09810

(43) International Publication Date: 15 December 1988 (15.12.88)

(21) International Application Number: PCT/US88/02009

(22) International Filing Date: 11 June 1988 (11.06.88)

(31) Priority Application Number: 061,874

(32) Priority Date: 11 June 1987 (11.06.87)

(33) Priority Country: US

(71) Applicant: SYNTHETIC GENETICS [US/US]; 10457 Roselle Street, Suite E, San Diego, CA 92121 (US).

(72) Inventor: TULLIS, Richard, H. ; 1320 Saxony Street, Leucadia, CA 92024 (US).

(74) Agent: ROWLAND, Bertram, I.; 350 Cambridge Avenue, Suite 200, Palo Alto, CA 94306 (US).

(81) Designated States: AT (European patent), BE (European patent), CH (European patent), DE (European patent), FR (European patent), GB (European patent), IT (European patent), JP, LU (European patent), NL (European patent), SE (European patent).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: NOVEL AMPHIPHILIC NUCLEIC ACID CONJUGATES

(57) Abstract

Novel oligonucleotide conjugates are provided, where oligonucleotides are joined through a linking arm to a hydrophobic moiety. The resulting conjugates are more efficient in membrane transport, so as to be capable of crossing the membrane and effectively modulating a transcriptional system. In this way, the compositions can be used *in vitro* and *in vivo*, for studying cellular processes, protecting mammalian hosts from pathogens, and the like.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	FR	France	ML	Mali
AU	Australia	GA	Gabon	MR	Mauritania
BB	Barbados	GB	United Kingdom	MW	Malawi
BE	Belgium	HU	Hungary	NL	Netherlands
BG	Bulgaria	IT	Italy	NO	Norway
BJ	Benin	JP	Japan	RO	Romania
BR	Brazil	KP	Democratic People's Republic of Korea	SD	Sudan
CF	Central African Republic			SE	Sweden
CG	Congo	KR	Republic of Korea	SN	Senegal
CH	Switzerland	LI	Liechtenstein	SU	Soviet Union
CM	Cameroon	LK	Sri Lanka	TD	Chad
DE	Germany, Federal Republic of	LU	Luxembourg	TG	Togo
DK	Denmark	MC	Monaco	US	United States of America
FI	Finland	MG	Madagascar		

NOVEL AMPHIPHILIC NUCLEIC ACID CONJUGATES

INTRODUCTION

Technical Field

The subject invention relates to specific polynucleotide binding polymers conjugated to solubility modifying moieties for inhibition of expression.

Background

There is a continuing interest and need for agents capable of modulating intracellular expression. The agents could have a profound capability of solving a variety of genetically associated problems. These agents, particularly complementary nucleic acid agents, could be used as antiviral agents to inhibit the expression of viral essential genes. The agents also could act as anti-neoplastic agents, reducing the rate of proliferation of cancer cells or inhibiting their growth entirely. These agents would act intracellularly binding to transcription products by a mechanism or mechanisms unknown, to inhibit the expression of a particular structural gene.

There has been substantial interest in this possibility and a number of experiments in culture have shown that there may be some promise to this approach. However, there are also numerous short-comings to the approaches that have been used previously. In order to provide for a useful agent for therapy, the agent should be effective at low concentrations, so as to allow for relatively low dosages when administered systemically. Secondly, agents should be relatively stable and resistant to degradation by the various nucleases. Thirdly, the agent should be very rapid once introduced into the cytoplasm and highly specific in

binding to its complementary sequence, so as to avoid long incubation periods. Fourth, the agent should be able to penetrate the membrane. The agent should be effective at low concentrations to avoid high concentrations in the blood stream. Finally, adverse effects to the mammalian host should be minimized and the oligonucleotide agent should provide for a minimal immunogenic response. While various of these criteria may be compromised to different degrees, the agents which have been produced so far fall far short of agents which might find general use.

Relevant Literature

Use of relatively short probes to maximize selectivity while retaining high sensitivity to single base mismatches is suggested by Szostak, et al., Methods Enzymol. (1979) 68:419-429; Wu, Nature New Biology (1972) 236:198; Itakura and Riggs, Science (1980) 209:1401; Noyes, J. Biol. Chem. (1979) 254:7472-7475; Noyes et al., Proc. Natl. Acad. Sci. USA (1979) 76:1770-1774; Agarwal, et al., J. Biol. Chem. (1981) 256:1023-1028. Tullis, et al., Biochem. Biophys. Res. Comm. (1980) 93:941; Orkin et al., J. Clin. Invst. (1983) 71:775; Conner et al., Proc. Natl. Acad. Sci. USA (1983) 80:278; Piratsu et al., New Eng. J. Med. (1983) 309:284-287; Wallace et al., Gene (1981) 16:21.

There have been a number of reports on the use of specific nucleic sequences to inhibit viral replication. See for example, Zamecink and Stephenson, Proc. Natl. Acad. Sci. USA (1978) 75:280-284; Tullis et al., J. Cellular Biochem. Suppl. (1984) 8A:58 (Abstract); Kawasaki, Nucl. Acids. Res. (1985) 13:4991; Walder et al., Science (1986) 233:569-571; Zamecnik et al., Proc. Nat'l. Acad. Sci., USA (1986) 83:4143-4146.

Modified nucleic acids, such as triesters and methylphosphonates have also been shown to be effective in inhibiting expression. Miller et al., Biochemistry

(1974) 13:4887-4895; Barrett et al., Ibid. (1974) 13:4897-4906; Miller et al., Ibid. (1977) 16:1988-1997; Miller et al., Biochemistry (1981) 20:1873-1880; Blake et al., Biochemistry (1985a, b) 24:6132 and 6134; Smith et al., Proc. Nat'l. Acad. Sci. USA (1986) 83:2787-91; Agris et al., Biochemistry (1986) 25:6268-6275; Miller et al., Biochemistry (1986) 25:5092-5097.

Modified nucleic acid sequences for enhancing binding to the complementary sequence are reported by Vlassov et al., Adv. Eng. Reg. 1986):301-320; Summerton J. Theor. Biol. (1979) 78:77-99; Knorre (1986) Adv. Eng. Reg. 1986:277-300.

Reduced immunogenicity of proteins conjugated to polyethyleneglycol is report by Tomasi and Fallow, WO86/04145 (PCT/U585/02572) and Abuchowski et al., Cancer Biochem. Biophys. (1984) 7:175-186. See also U.S. Patent Nos. 4,511,713 and 4,587,044.

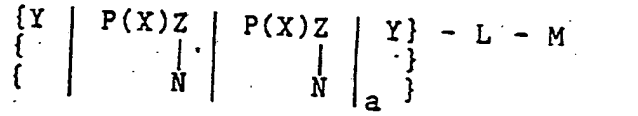
SUMMARY OF THE INVENTION

Novel nucleic acid conjugates are provided comprising a relatively short nucleic acid sequence complementary to a sequence of interest for modifying intracellular expression, a linking group, and a group which imparts amphiphilic character to the final product, usually more hydrophobic than hydorphilic, where hydrophobic includes amphiphilic. The nucleic acid moiety may include normal or other sugars, phosphate groups or modified phosphate groups or bases other than the normal bases where the modifications do not interfere with complementary binding of the sequence of interest. The compositions find use for inhibiting mRNA maturation and/or expression of particular structural genes, such as in neoplastic cells, of viral proteins in viral infected cells, and essential protein(s) of human and animal pathogens.

DESCRIPTION OF THE SPECIFIC EMBODIMENTS

The subject invention provides novel nucleic acid conjugates for inhibiting intracellular mRNA maturation and/or expression of a structural gene. Conjugates comprise a relatively short oligonucleotide sequence, a linking group, and a group which modifies the HLB (hydrophilic lipophilic balance) to provide an amphiphilic product product. The amphiphilic nature of the product aids in the transport of the conjugate across the cellular membrane and can provide additional advantages, such as increasing aqueous or liquid solubility of nucleic acid derivatives, e.g., use of an amphiphilic group to enhance water solubility of long chain methyl phosphonates and stabilizing normal nucleic acids to exonuclease digestion.

For the most part, compounds of this invention will have the following formula:



X is usually a pair of electrons, chalcogen (oxygen or sulfur) or amino, particularly NH;

Z is a naturally occurring or synthetic sugar residue linked at two of the 2', 3' and 5' hydroxyls of the five carbon sugars and at comparable sites for six carbon sugars, where the sugars will usually be ribose, or deoxyribose, or other 5 carbon or 6 carbon, particularly 5 carbon, sugars such as arabinose, xylose, glucose, or galactose;

N is any natural or unnatural base (purine or pyrimidine) capable of binding to and hybridizing with a natural purine or pyrimidine, the purines and pyrimidines may be the natural deoxyribose nucleoside purines and pyrimidines, such as adenine, cytidine, thymidine, guanidine or other purines and pyrimidines, such as uracil, inosine, and the like.

L is a linking group which is derived from a polyvalent functional group having at least 1 atom, not more than about 60 atoms other than hydrogen, usually not more than about 30 atoms other than hydrogen, having up to about 30 carbon atoms, usually not more than about 20 carbon atoms, and up to about 10 heteroatoms, more usually up to about 6 heteroatoms, particularly chalcogen, nitrogen, phosphorous, etc., non-oxocarbonyl (carboxy carbonyl), oxo-carbonyl (aldehyde or ketone), or the sulfur or nitrogen equivalents thereof, e.g., thiono, thio, imidyl, etc. as well as disulfide, amino, diazo, hydrazino, oximino, etc., phosphate, phosphono, and the like.

M is a solubility modifying moiety which imparts amphiphilic character to the molecule, particularly hydrophobic with phosphates and amphiphilic with phosphonates, which will have a ratio of carbon to heteroatom of at least 2:1, usually at least 3:1, frequently up to greater than 20:1, may include hydrocarbons of at least 6 carbon atoms and not more than about 30 carbon atoms, polyoxy compounds (alkyleneoxy compounds), where the oxygen atoms are joined by from about 2 to 10 carbon atoms, usually 2 to 6 carbon atoms, preferably 2 to 3 carbon atoms, and there will be at least about 6 units and usually not more than about 200 alkyleneoxy units, more usually not more than about 100 units, and preferably not more than about 60 units.

One Y is a bond to L, while the other Y is a monovalent oxy, thio, amino, sugar group or substituted functionalities thereof, or alkyl of up to about 20, usually of up to about 6 carbon atoms, when bonded to P, or hydrogen, hydrocarbyl or acyl of from 1 to 30, usually 1 to 12 carbon atoms, or substituted hydrocarbyl or acyl having from 1 to 4 hetero groups which are oxy, thio, or amino when bonded to Z.

a is at least 5 and not more than about 50, usually not more than about 35.

The phosphorous moiety may include phosphate, phosphoramidate, phosphordiamidate, phosphorothioate, phosphorothionate, phosphorothiolate, phosphoramidothiolate, phosphonate, phosphorimidate and the like.

The purines and pyrimidines may include thymidine, uracil, cytosine, 6-methyluracil, 4,6-dihydroxypyrimidine, isocytosine, hypoxanthine, xanthine, adenosine, guanosine, and the like.

The sugars may be ribose, arabinose, xylylose or α-deoxy derivatives thereof. Other nucleosides may also employ hexoses.

A wide variety of linking groups may be employed, depending upon the nature of the terminal nucleotide, the functionality selected for, whether the linking group is present during the synthesis of the oligonucleotide, the functionality present on the solubility modifying moiety and the like. A number of linking groups are commercially available and have found extensive use for linking polyfunctional compounds. The linking groups include: $-OCH_2CH_2NHCO(CH_2)_nCONH-$; $-OCH_2CH_2NH-X-(CH_2)_nNH-$;$-O-P(O)(OH)NHCO(CH_2)_nCONH-$; $OCH_2CH_2NHCO\phi S-$;$-NH(CH_2)_nNH$;$-O(CH_2)_nO-$;$-O(CH_2CH_2NH)_m-$; $-NH(CH_2)_nSYN$; $-CO(CH_2)_nCO$; $-SCH_2CH_2CO-$; $-CO\phi NYS-$; $-(NCH_2CH_2)_mCH_2N-$; $-O(CO)NH(CH_2)nNH$; charged and uncharged homo- and copolymers of amino acids, such as polyglycine, polylysine, polymethionine, etc. usually of about 500 to 2,000 daltons; wherein X is 2,5-quinondiyl, Y is (3-succindioyl) to form succinimidyl, n is usually in the range of 2 to 20, more usually 2 to 12, and m is 1 to 10, usually 1 to 6.

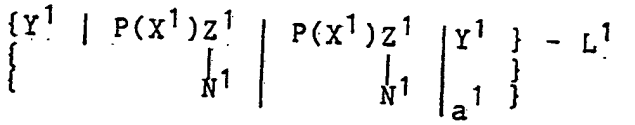
The lipophilic/amphiphilic group may be a wide variety of groups, being aliphatic, aromatic, alicyclic, heterocyclic, or combinations thereof, usually of at least 6, more usually at least 12 and not more than

about 500, usually not more than about 200 carbon atoms, having not more than about 1 heteroatom per 2 carbon atoms, being charged or uncharged, including alkyl of at least 6 carbon atoms and up to about 30 carbon atoms, usually not more than about 24 carbon atoms, fatty acids of at least about 6 carbon atoms, usually at least about 12 carbon atoms and up to about 24 carbon atoms, glycerides, where the fatty acids will generally range from about 12-24 carbon atoms, there being from 1-2 fatty acids, usually at the 2 or 3 positions or both, aromatic compounds having from 1 to 4 rings, either mono- or polycyclic, fused or unfused, polyalkyleneglycols where the alkylenes are of from 2-10, usually of from 2-6 carbon atoms, more usually 2-3 carbon atoms, there being usually at least about 6 units, more usually at least about 10 units, and usually fewer than about 500 units, more usually fewer than about 200 units, preferably fewer than about 100 units, where the alkylene glycols may be homopolymers or copolymers; alkylbenzoyl, where the alkyl group will be at least about 6 carbon atoms, usually at least about 10 carbon atoms, and not more than about 24 carbon atoms, usually not more than about 20 carbon atoms; alkyl phosphates or phosphonates, where the alkyl group will be at least about 6 carbon atoms, usually at least about 12 carbon atoms and not more than about 24 carbon atoms, usually not more than about 20 carbon atoms, or the like.

The solubility modifying group may be charged or uncharged, preferably being uncharged, under physiological conditions, usually having not more than 1 charge per 10 atoms of the group other than hydrogen. Illustrative groups include polyethylene glycol having from about 40-50 units, copolymers of ethylene and propylene glycol, laurate esters of polyethylene glycols, triphenylmethyl, naphthylphenylmethyl, palmitate, distearylglyceride, didodecylphosphatidyl, cholesteryl, arachidonyl, octadecanyloxy, tetradecylthio, etc.

Functionalities which may be present include oxy, thio, carbonyl (oxo or non-oxo), cyano, halo, nitro, aliphatic unsaturation, etc.

Of particular interest will be oligonucleotide conjugates of the following formula:



X^1 is nitrogen or oxygen;

Z^1 is ribose or deoxyribose substituted at the 3' and 5' positions;

One Y^1 is a bond to L^1 and the other Y^1 is hydroxy, alkyl, alkoxy or amino (including substituted amino, e.g., alkyl, acyl, etc.) of from 0 to 3 carbon atoms or a five carbon sugar, particularly ribose or deoxyribose, when bonded to P and hydrogen, alkyl, or acyl of from 1 to 10, usually 1 to 6 carbon atoms when bonded to Z^1;

N^1 is any purine or pyrimidine which can hybridize to the naturally occurring purines and pyrimidines, but is preferably a naturally occurring purine or pyrimidine;

L^1 is a linking group of at least about 2 carbon atoms and not more than about 30 carbon atoms, usually not more than about 20 carbon atoms, having from 0-10, usually 1-6 heteroatoms, which will be oxygen, nitrogen, and sulfur, particularly as oxy, amino, or thio;

M^1 is the solubility modifying moiety, hydrophobic or amphiphilic, which is desirably a polyalkyleneoxy group of at least about 20 units and not more than about 200 units, normally not more than about 150 units, where the alkylene groups are of from 2-3 carbon atoms;

a^1 is at least 5, usually at least 7 and generally not more than about 50, usually not more than about 30, more usually ranging from about 11 to 30, preferably from about 13 to 30.

In preparing the subject compositions, the oligonucleotide and the solubility modifying moiety will usually exist as independent moieties and may be joined together by a linker arm. The oligonucleotide may be made by any convenient synthetic procedure. For the most part, recombinant procedures will not be employed, although in some situations they may be useful. Various commercial synthetic devices for preparing polynucleotides are available from a number of companies, such as Applied Biosystems Inc., Biosearch, Inc. and Pharmacia. A variety of procedures are known for employing blocked oligonucleotides as their triesters, phosphoramidites, phosphonates, or the like, where a cycling procedure is employed, and the individual nucleotides are added in succession.

At the completion of the synthesis, various protocols may be employed. Preferably in most cases, the terminal blocking group may be removed and the linking arm joined to the terminal nucleotide. Alternatively, all of the blocking groups may be removed and the terminal nucleotide modified, by addition of the linking arm, where the linking arm may be specific for the final oligonucleotide. In some instances, the terminal blocking group may serve as all or part of the linking arm. Alternatively, the oligonucleotide may be removed from the support and then manipulated further, particularly where the linking group to the support may be used as the linking arm for joining the hydrophobic modifying moiety. Various procedures for further functionalization of the 5'- or 3'-termini of oligonucleotides may be found in Chu and Orgel DNA (1985) 4:327-331; Connolly and Rider Nucl. Acids Res. (1985) 13: 4485-4502.

Depending upon the functionalities, various reactions may be employed to produce amides, esters, both inorganic and organic, oxygen and sulfur ethers, amines, or the like. In working with carboxyl groups, various activating groups may be employed, such as carbonyldiimidazole, carbodiimides, succinimidyl ester, para-nitrophenyl ester, etc.

Various active functionalities can be employed, such as isocyanates, diazo groups, imino chlorides, imino esters, anhydrides, acyl halides, sulfinyl halides, isothiocyanates, sulfonyl chlorides, etc. Conditions for carrying out the various reactions in joining non-nucleotide moieties to nucleotide moieties may be found in Chu and Orgel DNA (1985) 4:327-331; Smith, et al. Nucl. Acids. Res. (1985) 13:2399-2412.

The solubulity modifying moiety may be added to the linking arm either prior to, subsequent to or concurrently with the addition of the linking arm to the oligonucleotide. For the most part, the solubility modifying moiety will be added subsequent to the reaction of the linking arm to the oligonucleotide. In some instances, it may be desirable to join the solubility modifying moiety to the linking arm, where the linking arm is bound to the oligonucleotide while the oligonucleotide is still bound to the support. As already indicated, the reactions between the linking arm and the solubility modifying moiety will vary with the particular functional groups present, the nature of the hydrophobic moiety, reaction conditions which are required, and the like.

For the most part, reaction conditions will be mild, and will occur in polar solvents or combinations of polar and non-polar solvents. Solvents will vary and include water, acetonitrile, dimethylformamide, diethyl ether, methylene chloride, etc. Reaction temperatures will be for the most part in the range of about

-10 to 60°C. Usually, after completion of the reaction between components of the conjugate, the resulting product will be subjected to purification.

The manner of purification may vary, depending upon whether the oligonucleotide is bound to a support. For example, where the oligonucleotide is bound to a support, after addition of the linking arm to the oligonucleotide, unreacted chains may be degraded, so as to prevent their contaminating the resulting product. On such cases, the bonding of the linker to the oligonucleotide must be sufficiently stable to withstand the cleavage conditions from the synthesis support, e.g., conc. ammonia. Where the oligonucleotide is no longer bound to the support, whether only reacted with the linking arm or as the conjugate to the solubulity modifying moiety intermediate or as the final product, each of the intermediates or final product may be purified by conventional techniques, such as electrophoresis, solvent extraction, HPLC, chromatography, or the like. The purified product is then ready for use.

The subject products will be selected to have an oligonucleotide sequence complementary to a sequence of interest. The sequence of interest may be present in a prokaryotic or eukaryotic cell, a virus, a normal or neoplastic cell. The sequences may be bacterial sequences, plasmid sequences, viral sequences, chromosomal sequences, mitochondrial sequences, plastid sequences, etc. The sequences may involve open reading frames for coding proteins, ribosomal RNA, snRNA, hnRNA, introns, untranslated 5'- and 3'-sequences flanking open reading frames, etc. The subject sequences may therefore be involved in inhibiting the availability of an RNA transcript, inhibiting expression of a particular protein, enhancing the expression of a particular protein by inhibiting the expression of a repressor, reducing proliferation of viruses or neoplastic cells, etc.

The subject conjugates may be used in vitro or in vivo for modifying the phenotype of cells, limiting the proliferation of pathogens such as viruses, bacteria, protists, mycoplasma, chlamydia, or the like, or inducing morbidity in neoplastic cells or specific classes of normal cells. Thus, one can use the subject compositions in therapy, by administering to a host subject to or in a diseased state, one or more of the subject compositions to inhibit the transcription and/or expression of the native genes of the cell. The subject compositions may be used for protection from a variety of pathogens in a mammalian host, e.g., enterotoxigenic bacteria, Pneumococcus, Neisseira, etc.; protists, such as Giardia, Entamoeba, etc.; neoplastic cells, such as carcinoma, sarcoma, lymphoma, etc.; specific B-cells, specific T-cells, such as helper cells, supressor cells, CTL, NK, ADCC, etc.

The subject sequences may be selected so as to be capable of interfering with transcription product maturation or expression of proteins by any of the mechanisms involved with the binding of the subject composition to its target sequence. These mechanisms may include interference with processing, inhibition of transport across the nuclear membrane, cleavage by endonucleases, or the like.

The subject sequences may be complementary to such sequences as sequences expressing growth factors, lymphokines, immunoglobulins, T-cell receptor sites, MHC antigens, DNA or RNA polymerases, antibiotic resistance, multiple drug resistace (mdr), genes involved with metabolic processes, in the formation of amino acids, nucleic acids, or the like, DHFR, etc. as well as introns or flanking sequences associated with the open reading frames.

The following table is illustrative of some additional applications of the subject compositions.

THERAPEUTIC APPLICATIONS OF SYNTHETIC DNA TECHNOLOGY

Area of Application	Specific Application Targets
Infectious Diseases:	
Antivirals, Human	AIDS, Herpes, CMV
Antivirals, Animal	Chicken Infectious Bronchitis Pig Transmissible Gastroenteritis Virus
Antibacterial, Human	Drug Resistance Plasmids, E. coli
Antiparasitic Agents	Malaria Sleeping Sickness (Trypanosomes)
Cancer	
Direct Anti-Tumor Agents	c-myc oncogene - leukemia other oncogenes
Adjunctive Therapy	Methotrexate Resistance - leukemia Drug Resistant Tumors - drug transport
Auto Immune Diseases	
T-cell receptors	Rheumatoid Arthritis Type I Diabetes Systemic Lupus Multiple sclerosis
Organ Transplants	Kidney - OTK3 cells cause GVHD

The subject compositions may be administered to a host in a wide variety of ways, depending upon whether the compositions are used *in vitro* or *in vivo*. *In vitro*, the compositions may be introduced into the nutrient medium, so as to modulate expression of a particular gene by transfer across the membrane into the cell interior such as the cytoplasm and nucleus. The subject compositions can find particular use in protecting mammalian cells in culture from mycoplasma, for modifying phenotype for research purposes, for evaluating the effect of variation of expression on various metabolic processes, e.g., production of particular products, variation in product distribution, or the

like. While no particular additives are necessary for transport of the subject compositions intracellularly, the subject compositions may be modified by being encapsulated in liposomes or other vesicles, and may be used in conjunction with permeabilizing agents, e.g., non-ionic detergents, Sendai virus, etc.

For in vivo administration, depending upon its particular purpose, the subject compositions may be administered in a variety of ways, such as injection, infusion, tablet, etc., so that the compositions may be taken orally, intravascularly, intraperitoneally, subcutaneously, intralesionally, or the like. The compositions may be formulated in a variety of ways, being dispersed in various physiologically acceptable media, such as deionized water, water, phosphate buffered saline, ethanol, aqueous ethanol, or formulated in the lumen of vesicles, such as liposomes or albumin microspheres.

Because of a wide variety of applications and manners of administration, no particular composition can be suggested. Rather, as to each indication, the subject compositions may be tested in conventional ways and the appropriate concentrations determined empirically. Other additives may be included, such as stabilizers, buffers, additional drugs, detergents, excipients, etc. These additives are conventional, and would generally be present in less than about 5 wt%, usually less than 1 wt%, being present in an effective dosage, as appropriate. For fillers, these may be as high as 99.9% or greater of the composition, depending upon the amount of active material necessary.

The following examples are presented by way of illustration not by way of limitation.

EXPERIMENTAL

EXAMPLE 1

Synthesis of Polyethylene Glycol Derivatives of Normal DNAs Using Aminolink, Benzoquinone and Bis-(Aminohexyl) Polyethylene Glycol

Chemical Synthesis of DNA oligonucleotides by the Amidite Method.

The chemical synthesis of DNA can be carried out using slight modifications of the conventional phosphoramidite methods on any commercially available DNA synthesizer. This method is a modification of the technique described by Caruthers and coworkers (Beaucage and Caruthers, Eur. Pat. Appl. 82/102570.

In this technique, 0.1 M nucleoside phosphoramidites dissolved in anhydrous acetonitrile were mixed with an equal volume of 0.5 M tetrazole and sequentially coupled to the 5'-hydroxyl terminal nucleotide of the growing DNA chain bound to controlled pore glass supports via a succinate spacer (Matteucci and Caruthers, Tetrahedron Letters (1980) 21:719-22. Nucleoside addition was followed by capping of unreacted 5'-hydroxyls with acetic anhydride, iodine oxidation, and 5'-detritylation in trichloroacetic acid-methylene chloride. The resin-bound oligomer was then dried by extensive washing in anhydrous acetonitrile and the process repeated. Normal cycle times using this procedure were 12 minutes with condensation efficiencies of >98% (as judged by trityl release).

As the last step of the synthesis, trityl was removed from the product oligonucleotide chains and an aminoethanolphosphoramidite was added to the 5'-hydroxyl using Aminolink (Applied Biosystems, Foster City, CA). The resin-bound oligonucleotide was then deblocked and released from the column using a method appropriate to the type of phosphate linkage present.

For normal phosphodiesters, release from the column and hydrolysis overnight at 55°C in concentrated ammonium hydroxide was appropriate.

The product was then lyophilized several times from 50% aqueous ethanol and purified via reversed phase HPLC C-8 silica columns, eluting with 5 to 50% acetonitrile/25 mM ammonium acetate, pH 6.8 in a linear gradient. If required, the material may be further purified by ion-exchange HPLC on Nucleogen DEAE 60-7 eluting with 20% acetonitrile/25 mM ammonium acetate, pH 6.5. The recovered product was then characterized by gel electrophoresis on 15% polyacrylamide gels carried out as described by Maxam and Gilbert in Methods of Enzymology (1980) 68:499-560. Oligonucleotides in finished gels were visualized using Stains-all. The Stains-All procedure did not work for uncharged oligonucleotides such as DNA methylphosphates or ethyl triesters.

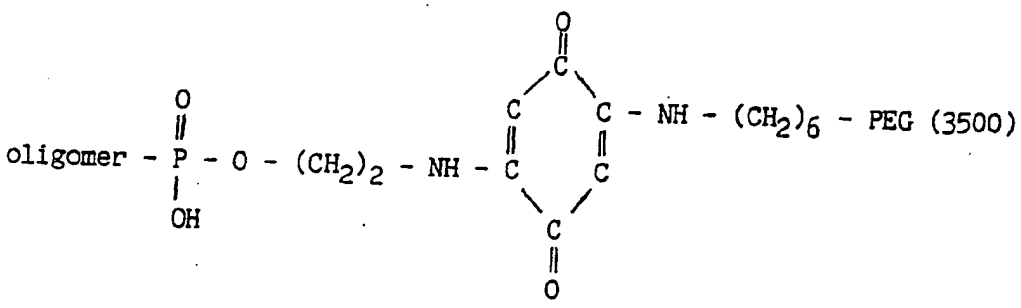
The fully deblocked and purified product is then converted to the appropriate polyethylene glycol derivative using a suitable coupling technique. Several techniques can be used including benzoquinone, carbodiimide, SMCC (Succinimidyl 4-(N-maleimideomethyl)-cyclohexane-1-carboxylate, SPDP (N-succinimidyl 3-(2-pyridyldithio)propionate, carbonyldiimidazole, Aminolink, disuccinimidyl suberimidate and phenylisocyanate.

Coupling of the linker arm DNA to benzoquinone and cross-linking to Bis(aminohexyl) polyethylene glycol.

In the first step bis-(aminohexyl)polyethylene glycol is reacted with a 100 to 1000 fold molear excess of benzoquinone in 0.1 M sodium bicarbonate (pH 8.5). After 1 hour at room temperature, the excess unreacted benzoquinone is removed by Sephadex G-25 column chromatography. The activated polyethylene glycol is then made to 0.1 M sodium bicarbonate and reacted with the DNA oligomer containing a reactive amine linker arm in

a molar ratio of 10:1 and the reaction allowed to proceed to completion. At the end of the reaction (generally overnight) the unreacted oligomer is removed by gel-filtration on Sephadex G-100 and the complex characterized by polyacrylamide gel electrophoresis (cf. Maniatis, et al., Molecular cloning, A laboratory manual (1982) Cold Spring Harbor Laboratories, Cold Spring Harbor, NY). Further purification can be effected using ion-exchange chromatography and gel electrophoresis as required.

The structure of the product of these reactions is:



EXAMPLE 2

Synthesis of Polyethylene Glycol Derivatives of Normal DNAs Using Aminolink and Carbonyldiimidazole Activated Polyethylene Glycol

In this example the Aminolink oligonucleotide was synthesized as described in Example I. After removal of the oligomer from the support and deblocking in ammonia, the solution was evaporated in vacuo and dissolved in 0.1M NaHCO3, pH 8.5 and purified on a G25-spun column to convert the material to the sodium salt and to remove any extraneous amine-containing material of low molecular weight. The solution was then made to 0.2 M in carbonyldiimidazole-activated polyethylene glycol (MW_{av} = 20,000) and allowed to react overnight at 23°C.

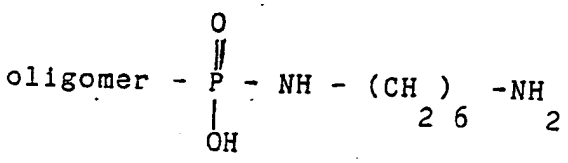
Unbound oligonucleotide was removed by gel filtration on Sephadex G-100. On this column the complex elated in the excluded volume of the column while the free oligonucleotide and unbound polyethylene glycol were retained. This material was then concentrated *in vacuo* and the complex characterized by polyacrylamide gel electrophoresis (Maniatis *et al.*, (1982), *supra*.

EXAMPLE 3

Synthesis of Polyethylene Glycol Derivatives of Normal DNAs Using Phosphoramidate Linker Amines and N-Hydroxysuccinimidyl Activated Polyethylene Glycol

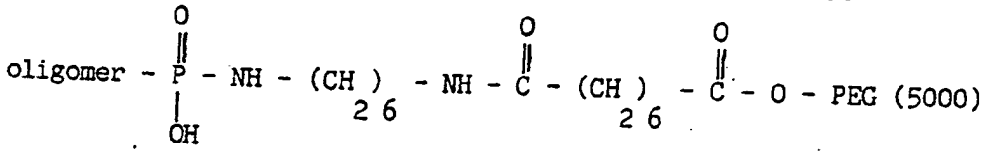
In this method DNA is synthesized as in Example 1 with the exception that the trityl group is removed without the further addition of the Aminolink phosphoramidite. After purification by polyacrylamide gel electrophoresis, the product DNA is phosphorylated with the forward reaction of T4 polynucleotides kinase according to standard procedures (Miller *et al.*, *Nucl. Acids. Res.* (1983) 11:6225-42; Maniatis *et al.*, (1982), *supra*; Maxam and Gilbert, *Proc. Nat'l Acad. Sci. USA* (1980) 74:560-5. Labeled oligomers can be separated from unreacted ATP by DEAE chromatography and C-18 reverse phase columns (e.g. Waters C-18 SepPak). Samples are checked for purity on analytical 20% polyacrylamide gels.

The phosphorylated oligomer is then treated with 1-methylimidazole and hexanediamine, in the presence of EDC carbodiimide according to the method of Chu and Orgel *DNA* (1985) 4:327-31. This reaction covalently couples the diamine linker to the oligonucleotides via a phosphoramidate linkage with the following structure:



The amine linker arm oligomer is then conjugated to NHS-succinylmonomethoxypolyethylene glycol (MW 5000) as follows. The oligonucleotide is dissolved to a final concentration of 100 μM per liter in 50 mM sodium phosphate buffer, pH 7.1 containing 0.15M NaCl. To this solution a 10 fold molar excess of SS-PEG (5000) is added as a dry solid, allowed to dissolve and the reaction mixture incubated overnight at 25°C. The product is then purified by gel filtration chromatography on Sephadex G-100 in water and characterized by polyacrylamide gel electrophoresis.

The structure of the final product is:



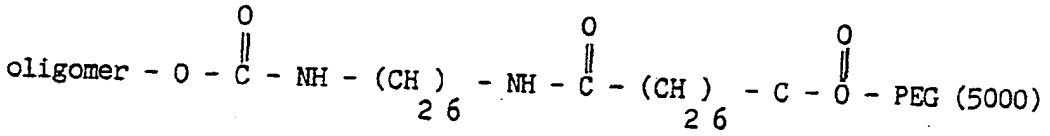
EXAMPLE 4

Synthesis of Polyethylene Glycol Derivatives of Normal DNAs Using Imidazole Activated Carboxylic Acid Esters and Bis-Aminoalkyl Polyethylene Glycol

In this example, DNA was synthesized according to the method given in Example 1. After synthesis, the product material was retained on the synthesis support with trityl removed from the 5' end of the molecule. The solid material was then thoroughly washed with anhydrous acetonitrile and blown dry under a stream of dry argon. Using a plastic syringe, 1 cc of 0.3M carbonyldiimidazole dissolved in anhydrous acetonitrile was pushed slowly through the synthesis column containing the support bound oligomer over the course of 1 hour. The 5' carbonylimidazole activated oligomer on the column was then washed free of excess reagent with 15 ml of acetonitrile and subsequently treated for 16

hours with 0.1 M bis (aminohexyl) polyethylene glycol in acetonitrile, water, acetonitrile and methylene chloride in succession. The polyethylene oligomer conjugate was then eluted with concentrated ammonium hydroxide and deblocked in the same by incubation at 55°C for 5 hours.

The reaction product is then purified by high performance gel filtration chromatography (HPGFC) on a TSK G4000SW column eluting 10mM Tris, pH 7.5 at 0.5 ml per minute. Further purification may be effected by agarose gel electrophoresis. The structure of the final conjugate synthesized by this method is:



EXAMPLE 5

Synthesis of Long Chain Alkane Derivatives of Normal DNAs Using Imidazole Activated Carboxylic Acid Esters and Aminoalkanes

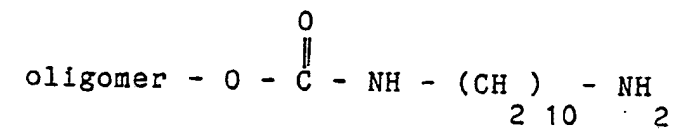
In this example, a 20 nucleotide DNA complementary to the initiation region of mouse β-globin mRNA was synthesized according to the method given in Example 1. After synthesis, the product material was retained on the synthesis support with trityl removed from the 5' end of the molecule. The solid material was then thoroughly washed with anhydrous acetonitrile and blown dry under a stream of dry argon. Using a plastic syringe, 1 cc of 0.3M carbonyldiimidazole dissolved in anhydrous acetonitrile was pushed slowly through the synthesis column containing the support-bound oligomer for 45 minutes. The 5' carbonylimidazole activated oligomer on the column was then washed free of excess reagent with 15 ml of acetonitrile and then treated with 0.2 M decanediamine in acetonitrile: water (10:1) for 30 minutes.

The material on the column was washed free of unreacted decanediamine with acetonitrile and water and then eluted from the column in concentrated ammonium hydroxide solution. After removal from the column, the ammonium hydroxide solution containing the oligomer conjugate was placed in a sealed vial and incubated 5 hours at 55°C.

The product was then lyophilized several times from 50% aqueous ethanol and purified via reversed phase HPLC C-8 silica columns eluted with 5 to 50% acetonitrile/25mM ammonium acetate, pH 6.8 in a linear gradient. If required, the material may be further purified by ion-exchange HPLC on Nucleogen DEAE 60-7 using 20% acetonitrile/25 mM ammonium acetate, pH 6.5 as eluent. The recovered product was then characterized by gel electrophoresis in 15% polyacrylamide gels carried out as described by Maxam and Gilbert in Meth. Enzymol. (1980) 68:499-560. Oligonucleotides in finished gels were visualized using Stains-all.

The presence of a primary amine was determined by two methods. First, reaction with fluorescamine produced a fluorescent product characteristic of the presence of a primary amine while no fluorescence was observed with similarly treated control oligomers of the same type but lacking the amine linker. Second, the decane conjugate was dissolved in 100 µl 0.1 M sodium bicarbonate to which was added 1 mg of fluoresceinisothiocyanate (FITC). After 1 hour of incubation, the unreacted FITC was removed by gel filtration chromatography on Sephadex G-25 spun columns. The product was then analysed by polyacrylamide gel electrophoresis as described above and the fluorescent band product visualized under UV illumination. A single fluorescent band was observed which corresponded to the oligomer visualized by subsequent staining with Stains-all.

The product of this reaction is an alkyl carbamate which is stable to moderate exposure to concentrated base. The structure of the final conjugate synthesized by this method is:



Other monoaminoalkyl and aryl derivatives can be produced by this method. Other molecules in this series which have been constructed include the derivatives made with ethylene diamine and hexane diamine. Higher chain length additions may require slight modifications of the solvent polarity in order to achieve the necessary concentrations. Alternatively, lower concentrations may be used if the reaction times are extended.

EXAMPLE 6

Synthesis of Polyethylene Glycol Derivatives of DNAs Using Imidazole-Activated Carboxylic Acid Esters, Polylysine Linker, DSS AND BIS-Aminoalkyl Polyethylene Glycol

In this example, a 25 nucleotide DNA complementary to the initiation region of mouse ß-globin mRNA was synthesized according to the method given in Example 1. After synthesis, the synthesis support was treated with 80% acetic acid for 30 minutes to remove trityl from the 5' end of the molecule. The solid material was then thoroughly washed with anhydrous acetonitrile and blown dry under a stream of dry argon and treated with 0.3M carbonyldiimidazole as in Example 4. The 5' carbonyldiimidazole-activated oligomer on the column was then washed free of excess reagent with 15 ml of acetonitrile and then treated with 0.2M poly-L-lysine (MW=1000) dissolved in 50% acetonitrile containing 0.1M sodium phosphate, pH 8 for 16 hours at room temperature.

The material on the column was washed free of salts and unreacted polylysine with water and acetonitrile and then eluted from the column with concentrated ammonium hydroxide. After removal from the column, the ammonium hydroxide solution containing the oligomer conjugate was incubated 5 hours at 55°C in a sealed glass vial. The product was then lyophilized several times from 50% aqueous ethanol and purified via gel filtration chromatography on TSK G4000SW in 10 mM Tris buffer, pH 7.5. The presence of a primary amine was determined by reaction with fluorescamine. No fluorescence was observed with control oligomers lacking the polyamine linker.

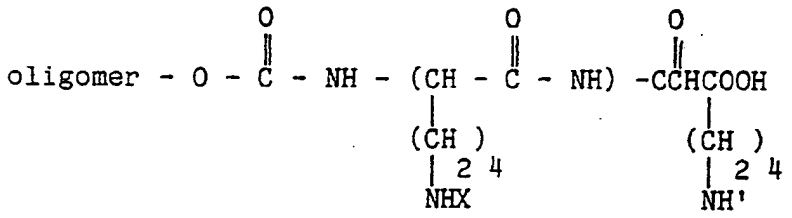
In order to render the polyamine conjugate negatively charged, the complex was reacted with FITC to label the molecule and to neutralize the positive charges on the amines. This was accomplished by dissolving a portion of the material in 100 µl 0.1M sodium bicarbonate to which was added 1 mg of FITC. After 1 hour of incubation, the unreacted FITC was removed by gel filtration chromatography on Sephadex G-25 spun columns (Maniatis et al., (1982), supra. The product was then analysed by polyacrylamide gel electrophoresis carred out as described by Maxam and Gilbert (1980) supra and the fluorescent band product visualized under UV illumination. A broad fluorescent band was observed which corresponds to the DNA visualized by Stains-all.

The oligomer containing polylysine covalently linked to the 5' end of the molecule was then cross-linked to bis-(aminohexyl) polyethylene glycol (MW = 3500) as follows. The polylysine oligomer is first dialysed against 0.1 M sodium carbonate, 3M NaCl and concentrated to a final concentration of 4 mg/ml using a Centricon 10 apparatus (Amicon, Danvers, N.J.). To 50 µl of this solution was added 25 µl of disuccinimidyl suberate (DSS, 10 mg/ml in DMSO) and the mixture incubated 10 minutes at room temperature. The unre-

acted DSS was then quickly removed by gel filtration on Sephadex G25 and concentrated on Centricon 10 membranes. The solution was then made to 0.2M in bis-(aminohexyl) polyethylene glycol and incubated overnight at room temperature to form the final conjugate. Purification was effected on TSK G4000 SW columns operated as previously described.

This conjugate has the following general formula:

1. Formulation Type I



Where X is usually H, at least one X being $-CO(CH_2)_6COHN-PEG_{5000}$.

By varying the reaction excess or the molecular weight of the polyethylene glycol and the polylysine used it is possible to construct polymer conjugates with varying degrees of substitution size and charge. The ability to vary these properties of the complex make it possible to design the use of the compound in various applications.

EXAMPLE 7

Synthesis of Polyethylene Glycol Derivatives of DNA Methylphosphonates

The chemical synthesis of DNA methylphosphonates (MP) may be carried out using a modification of the phosphochloridite method of Letsinger (Letsinger et al., J. Amer. Chem. Soc. (1975) 97:3278; Letsinger and Lunsford, J. Amer. Chem. Soc. (1976) 98:3605-3661; Tanaka and Letsinger, Nucl. Acids. Res. (1982) 25:3249-

60. In this procedure, dried blocked nucleosides dissolved in anhydrous acetonitrile 2,6-lutidine, are activated in situ with a stoichiometric amount of methyl dichlorophosphine. The activated nucleoside monochloridites are then added sequentially to the 5' hydroxy terminal nucleotide of the growing DNA chain bound to controlled pore glass supports via a succinate spacer (Matteucci and Caruthers, Tetrahed. Lett. (1980) 21:719-722. Each addition is followed by capping of unreacted 5'-hydroxyls with acetic anhydride, iodine oxidation, and 5'-detritylation in 3% trichloroacetic acid-methylene chloride.

The resin-bound methylphosphonate oligomers are then dried by extensive washing in anhydrous acetonitrile and the process repeated. Normal cycle times using this procedure are 23 minutes with condensation efficiencies of >32% (as judged by trityl release). The ultimate base may be added as the cyanoethyl phosphotriester which yields, upon cleavage in base, a 5'-terminal phosphodiester. This step makes it possible to radiolabel the oligonucleotide, purify and sequence the product using gel electrophoresis at intermediate stages of preparation (Narang et al., Can. J. Biochem. (1975) 53:392-394. Miller et al., Nucl. Acids Res. (1983) 11:6225-6242.

An amine-terminated linker arm is then added as follows. Trityl is removed as before and the resin treated with 0.2M Aminolink (Applied Biosystems, Foster City, CA) dissolved in dry acetonitrile containing 0.2M dimethylaminopyridine for 5 minutes. The linker arm oligonucleotide is then oxidized in iodine and washed in acetonitrile as above. Capping with acetic anhydride is not performed since any deblocked primary amine would be modified to the base-stable acetamide and thus be unavailable for further reaction.

At the end of the synthesis, the amine terminated linker arm methylphosphonate oligomer is base

deblocked as follows. The resin containing the DNA is removed from the column and placed in a water jacketed column and incubated in 1-2 ml phenol:ethylene diamine (4:1) for 10 hours at 40°C. At the end of the incubation in phenol:ethylene diamine, the resin is washed free of the phenol reagent and base protecting groups released using methanol, water, methanol and methylene chloride in succession. After drying in a stream of nitrogen, the intact, base-deblocked chains are cleaved from the support using EDA:ethanol (1:1) or a brief treatment at room temperature with ammonium hydroxide.

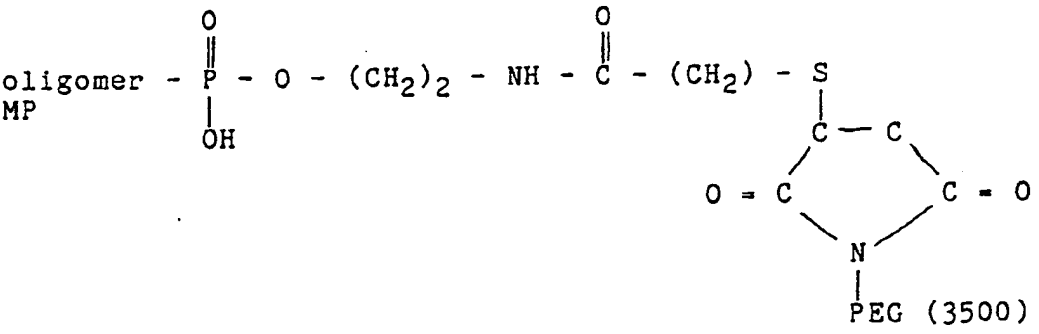
Purification of the amine-terminated DNA methylphosphonate is then performed as follows. The material is first lyophilized several times from 50% aqueous ethanol and purified via reversed phase HPLC C-8 silica columns eluted with 5 to 50% acetonitrile/ 25mM ammonium acetate, pH 6.8 in a linear gradient. Amine-containing fractions, as determined by fluorescamine reactivity, are pooled and the product recovered by drying in vacuo and further purified by ion-exchange HPLC on Nucleogen DEAE 60-7 eluted with 20% acetonitrile/ 25mM ammonium acetate, pH 6.5.

The purified product is then converted to the appropriate polyethylene glycol derivative using the heterobifunctional crosslinking agents SMCC and SATA (succinimidyl S-acetylthioacetate). Reactions using other reagents which can react with and modify the nucleoside bases (e.g. sulfonyl chlorides, glutaraldehyde or acid anydrides) are not recomended unless performed with the fully blocked oligonucleotide still bound to the synthesis support.

The DNA methylphosphonate containing 5' terminal reactive amine linker arms is first reacted with SATA in a 100-1000 fold molar excess at pH 8.5 (0.1M sodium bicarbonate). After 30 minutes at room temperature, the excess unreacted SATA is removed by G-25 column chromatography in water, concentrated in vacuo

and stored cold until ready for further reaction. Bis-(aminohexyl) polyethylene glycol is converted to the maleimide derivative by treatment with a 100-1000 fold molar excess of SMCC in 0.1M phosphate buffer, pH 6.9 for 1 hour at room temperature. Excess crosslinking agent is removed by chromatography on Sephadex G-100 and the material concentrated in vacuo and stored cold until ready for further reaction. This material is stable for about one week when kept cold. The SATA DNA methylphosphonate is then treated with hydroxylamine HCl dissolved in 0.1M phosphate buffer (pH adjusted to 7.2) for 1-2 hours. This treatment serves to release the reactive sulfhydryl. This product is then reacted overnight with a 10 fold molar excess of bis-(SMCC aminohexyl) polyethylene glycol by addition of the latter as a powder to the solution containing the oligomer.

Purification of the complex is then effected. Unbound oligonucleotide is removed by gel filtration on Sephadex G-100 or HPGFC on TSK G400SW eluted with 10mM Tris, pH 7.5. The diagrammatic structure of the final product of this procedure is:



EXAMPLE 8

Synthesis of Polyethylene Glycol Derivatives of DNA Alkyltriesters Using the Phosphoramidite Approach

The synthesis of the title compound triesters is performed according to the method of Zon and co-workers (Gallo et al., Nucl. Acids. Res. (1986)

14:7405-20; Summers et al., Nucl. Acids Res. (1986) 14:7421-36. The method of synthesis is similar to that used for in situ production of ethyl triesters as described in the other examples. Fully blocked dimethoxytrityl nucleosides are dried by repeated lyophilization from benzene, dissolved in anhydrous acetonitrile/2,6-lutidine and added dropwise to a stirred solution of chlorodiisopropylaminoethoxyphosphine in the same solvent at -70°C. The product is recovered by aqueous extraction, drying in vacuo and silica gel chromatography.

The chemical synthesis of DNA ethyl triesters (ETE) can be carried out using slight modifications of the conventional phosphoramidite methods. In this technique, nucleoside phosphoramidites dissolved in anhydrous acetonitrile are mixed with tetrazole and sequentially coupled to the 5'-hydroxy terminal nucleoside bound to CPG. Nucleoside addition is followed by capping of unreacted 5'-hydroxyls with acetic anhydride, iodine oxidation, and 5'-detritylation in trichloroacetic acid-methylene chloride. The resin-bound oligomer is then dried by extensive washing in anhydrous acetonitrile and the process repeated. Normal cycle times using this procedure are 17 minutes with condensation efficiencies of >96% (as judged by trityl release). The terminal residue is conventionally added as a diester in order to facilitate radiolabeling and purification. The 5'-terminal trityl group is left if HPLC purification is desired, but generally the 5'-terminal trityl is removed and the Aminolink procedure described in Example 1 is used.

At the end of the synthesis, the fully blocked product is base-deblocked as follows. The resin containing the fully protected DNA is removed from the column and placed in a water-jacketed chromatography column. The resin is then incubated in 1-2 ml phenol: ethylene diamine (4:1) for 10 hours at 40°C. At the

end of the incubation in phenol:ethylene diamine, the resin is washed free of the phenol reagent and base protecting groups released using methanol, water, methanol and methylene chloride in succession. After drying in a stream of nitrogen, the intact, base-deblocked chains are cleaved from the support using EDA:ethanol (1:1) or a brief treatment at room temperature with ammonium hydroxide.

Purification of the Aminolink DNA ethyl triester product is then performed as follows. The material is first lyophilized several times from 50% aqueous ethanol and purified via reversed phase HPLC C-8 silica columns eluted with 5 to 50% acetonitrile/25mM sodium acetate, pH 6.8 in a linear gradient. Amine-containing fractions as determined by fluorescamine reactivity are pooled and the product recovered by drying in vacuo and further purified by ion-exchange HPLC on Nucleogen DEAE 60-7 eluting 25% acetonitrile/25 mM ammonium acetate, pH 6.5.

The product oligonucleotide is then suitable for coupling to polyethylene glycol by any of the techniques previously described. In our experiments several techniques have been used, including SMCC, SPDP, carbonyldiimidazole, disuccinimidyl suberimidate and phenylisocyanate.

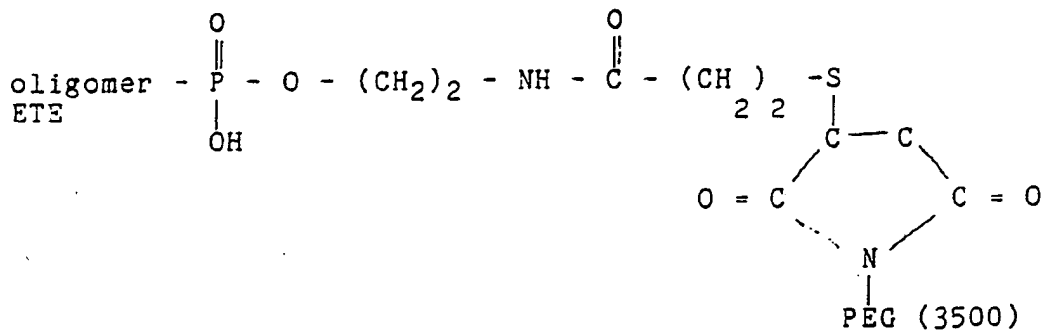
The SMCC/SPDP coupling reaction is as follows. The linker arm probe is coupled to excess SPDP followed by reduction with dithiothreitol (DTT), the unreacted DTT removed and the product allowed to cross-link through the free sulfhydryl to SMCC previously coupled to bis-(aminohexyl) polyethylene glycol (PEG). The formation of the thioether linkage is rapid and selective and the linkage formed is quite stable to a variety of conditions. The precise method of linkage formation is as follows:

The DNA containing amine linker arms is reacted with SPDP in a 100-1000 fold molar excess at pH

8.5 (0.1M sodium bicarbonate). After 1 hour at room temperature, the excess unreacted reagent is removed by G-25 column chromatography and the probe SPDP conjugate concentrated in vacuo. Bis-(aminohexyl) polyethylene glycol is converted to the maleimide derivative as described in the previous example. The SPDP DNA triester is then treated with 10 mM mercaptoethanol dissolved in 0.1M phosphate buffer (pH adjusted to 7.2) for 1 hour. This treatment serves to release the 5' thiopyridone thus forming a reactive sulfhydryl. Excess reducing agent is then removed using a G-25 spun column operated as previously described with the exception that all separations are performed in degassed 0.1M phosphate buffer, pH 6.8 under a nitrogen atmosphere to prevent the reoxidation of the terminal SH. In this procedure it is essential that all excess reducing agent be removed in order to prevent its subsequent reaction with the maleimidylated polyethylene glycol.

Thiopyridone released in this procedure provides a convenient indirect method for quantitating the presence of the 5'-terminal SH. Thiopyridone released by reduction has a UV absorption at 343nm. By following the increase in absorbance of the solution at this wavelength, the course of the reduction is easily followed. The thiopyridone can then be quantitated using a molar extinction coefficient of 8080. The product is then reacted overnight with a 10 fold molar excess of bis-(SMCC-aminohexyl) polyethylene glycol by addition of the latter as a powder or a concentrated solution to the solution containing the SH terminated oligomer triester. The reaction is allowed to proceed overnight at 25°C.

Purification of the complex is then effected. Unbound oligonucleotide is removed by gel filtration on Sephadex G-100 or HPGFC on TSK G4000SW eluted with 10mM Tris, pH 7.5. The diagrammatic structure of the final porduct of this procedure is:



EXAMPLE 9

Synthesis of Polyether Derivatives of DNA Alkyl and Aryltriesters Using the Phosphate Triester Approach

Synthesis of Phosphotriester Oligonucleotides of Varying Alkyl and Aryl Substituent Type.

The best available method for the production of the various triesters of variable alkane chain length is via conventional phosphate triester chemistry to synthesize the desired sequences as the b-chlorophenyl phosphate triesters (PTE). Upon completion of the synthesis, the fully protected oligonucleotide chlorophenyltriesters bound to the synthesis support are subjected to ester exchange in the presence of tetrabutylammonium fluoride and the desired alcohol. This basic method for the construction of DNA oligonucleotides is classical DNA synthesis chemistry. See Gait, (1984) Olignucleotide Synthesis: A Practical Approach, IRL Press, Washington, D.C.

The chemical synthesis of DNA p- or o-chlorophenyl phosphotriesters was carried out using a modification of the phosphochloridite method of Letsinger Tanaka and Letsinger, Nucl. Acids Res. (1982) 25:3249-60. For automated DNA synthesis, see Alvarado-Urbina et al., Science (1981) 214:270-273.

Fully blocked and dried nucleosides dissolved in anhydrous acetonitrile 2,6-lutidine and activated in situ with chlorophenoxydichlorophosphine are sequen-

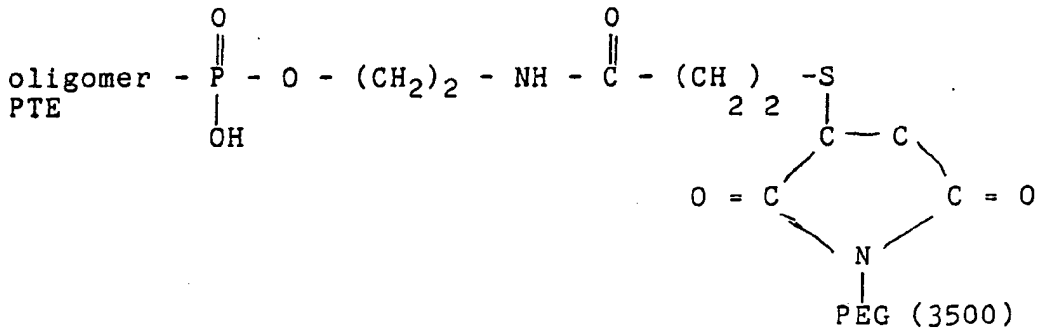
tially added to the 5'-hydroxy terminal nucleotide of the growing DNA chain bound to controlled pore glass supports via a succinate spacer as in previous examples. Derivatized glass supports, fully blocked nucleosides and other synthesis reagents are commercially available through Applied Biosystems (San Francisco, CA) or American Bionuclear (Emeryville, CA). Nucleoside addition is followed by capping of unreacted 5'-hydroxyls with acetic anhydride, iodine oxidation, and 5'-detritylation in trichloroacetic acid-methylene chloride.

The resin bound oligomer chlorophenyltriester is then dried by extensive washing in anhydrous acetonitrile and the process repeated. Normal cycle times using this procedure are 13 minutes with condensation efficiencies of >92% (as judged by trityl release). The ultimate base may be added as a β-cyanoethyl phosphotriester which yields, upon cleavage in base, a 5'-terminal phosphodiester. This step makes it possible to radiolabel the oligonucleotide and to purify and sequence the product using gel electrophoresis (Narang et al., Can. J. Biochem. (1975) 53:392-4; Miller et al., Biochemistry (1986) 25:5092-97.

The fully blocked material bound to the synthesis support is then subject to ester exchange in the presence of tetrabutylammonium fluoride (TBAF) and the desired alcohol under anhydrous conditions. This method yields rapid and quantitative alcohol exchange. The reaction is complete within 20 minutes for most aryl and alkyl alcohols which are capable of forming stable products.

In this example, anhydrous n-propanol is used to dissolve TBAF to a final concentration of 0.2M. The solution is then percolated slowly over the resin containing the oligomer chlorophenyl triester and allowed to react for about 1 hour at room temperature. The resin is then washed with methanol and acetonitrile and

dried under a stream of dry argon. Amine linker arm addition, deblocking and purification are then effected as in Example 8. Polyethylene glycol conjugation is performed as in Example 7. The final yield of conjugate is about 10% of the starting equivalents of nucleoside resin used. The diagrammatic structure of the final product is:



EXAMPLE 10

The Effect of Trityl Terminated Oligonucleotides on the Synthesis of β-globin Protein in vitro and in Cultured Cells

Using the methods of synthesis provided in the previous examples, both normal and ethyl triester type oligonucleotides were constructed. In the simplest example of an amphiphilic DNA conjugate containing a hydrophobic grouping at the 5' end of the molecule, the trityl group is left on at the end of the synthesis. Purified materials of this type were tested for their effectiveness in preventing the specific expression of hemoglobin in mouse erythroleukemia cells induced to produce hemoglobin. The oligonucleotides tested in these and the following examples are given in Table I.

TABLE I

DNA SEQUENCES SYNTHESIZED AND CONJUGATED FOR USE IN CELL CULTURE EXPERIMENTS

Probes Synthesized Antisense to Mouse Beta-globin mRNA	% GC	Sequence (3' to 5')
MBG 15 antisense	60%	G TAC CAC GTG GAC TG
MBG 15 antisense-DMT	60%	G TAC CAC GTG GAC TG-DMT
MBG 15 antisense-C_2 amine	60%	G TAC CAC GTG GAC TGp-O-$(CH_2)_2$-NH_2
MBG 15 ethyl triester	60%	g tac cac gtg gac tG
MBG 15 ethyl triester-DMT	60%	g tac cac gtg gac tG-DMT
MBG 20 antisense	55%	G TAC CAC GTG GAC TGA CTA C
MBG 20 antisense C_2	55%	G TAC CAC GTG GAC TGA CTA Cp-O-$(CH_2)_2$-NH_2
MBG 20 antisense C_6	55%	G TAC CAC GTG GAC TGA CTA C-O-(C))-NH-$(CH_2)_6$-NH_2
MBG 20 antisense C_{10}	55%	G TAC CAC GTG GAC TGA CTA C-O-(CO)-NH-$(CH_2)_{10}$-NH_2
MBG 20 antisense C_2-FITC	55%	G TAC CAC GTG GAC TGA CTA Cp-O-$(CH_2)_2$-NH_2-FITC
MBG 20 antisense C_6-FITC	55%	G TAC CAC GTG GAC TGA CTA C-O-(CO)-NH-$(CH_2)_6$-NH-FITC
MBG 20 antisense C_{10}-FITC	55%	G TAC CAC GTG GAC TGA CTA C-O-(CO)-NH-$(CH_2)_{10}$-NH-FITC
MBG 20 antisense C_2-PEG	55%	G TAC CAC GTG GAC TGA CTA Cp-O-$(CH_2)_2$-NH_2-PEG

a) Lower case letters represent nucleosides coupled to the 3' adjacent nucleoside via an ethyl phosphotriester linkage. Upper case letters represent 3' adjacent normal phosphodiester linkage. DMT represents a 5' terminal dimethoxytrityl moiety. C_2 derivatives are formed from the condensation of ethanolamine with a 5' terminal phosphate via an ester linkage. C_6 and C_{10} derivatives are the corresponding diamines coupled via an alkyl carbamate linkage to the 5' terminal hydroxyl. FITC represents the condensation product of fluoresceinisothiocyanate (isomer I) with the indicated diamine. PEG is polyethylene glycol (M_r = 3500).

The cells chosen for these experiments are Friend murine erythroleukemia (MEL) cells which can be induced to synthesize hemoglobin by a variety of agents including DMSO and butyric acid (cf. Gusella and Houseman, Cell (1976) 8:263-269. MEL cells are grown in culture using conventional techniques in a CO_2 incubator.

Induced cells which are expressing globin can be visualized by benzidine treatment which stains hemoglobin-producing cells blue (Leder et al., Science (1975) 190:893. Cells were exposed to the selected oligonucleotide conjugates at concentrations ranging from 1 mg/ml to 1 µg/ml during induction. Controls included mock-treated cells and cells treated with random sequence oligomer controls. Treated cells were scored at various time intervals for globin production based on staining intensity and the results compared to controls. About 50% of the control cells are inducible. Cell death or damage due to treatment is scored by Trypan blue exclusion in order to obtain an indication of toxicity and cell damage.

The results obtained are presented in Table II. These results show that the trityl terminated oligomers are more effective in producing the desired degree of synthesis inhibition. The trityl modified oligomers however showed some degree of cell damage which would not recommend their general use as therapeutic agents.

TABLE II

EFFECT OF TRITYLATED OLIGOMERS ON HEMOGLOBIN ACCUMULATION IN MOUSE CELLS

Oligomer Conjugate*	Viable Cells (% of Control)	% Benzidine* (B*)	% Inhibition B* Cells
DMSO Control	100%	100%	0%
MBG 15 100)M	100%	68%	32%
MBG 15 ETE 50)M	95%	59%	41%
MBG 15 ETE-DMT 50)M	94%	43%	57%

* See Table I. ETE is ethyl triester.

EXAMPLE 11

The Effect of Long Chain Alkyl Terminated Oligonucleotides on the Synthesis of β-globin Protein in Cultured Cells

Using the method of synthesis provided in the previous examples, 15 to 20 base long oligonucleotides conjugated to a 5'-terminal aminoalkane were constructed as described in Example 5. Purified materials of this type were tested for their effectiveness in preventing the specific expression of hemoglobin in MEL cells induced to produce hemoglobin. The results are given in Table III. The protocol for the test is given in Example 10.

TABLE III

THE EFFECT OF INCREASING HYDROPHOBICITY ON THE EFFECTIVENESS OF OLIGONUCLEOTIDES IN PREVENTING HEMOGLOBIN SYNTHESIS IN CULTURED CELLS

Treatment		Viable Cells	Inhibition of Benzidine* Cells
DMSO Control		46%	0%
MBG-20 Antisense	50 μM	50%	41%
MBG-20-C_2	50 μM	61%	41%
MBG-20-C_6	50 μM	60%	48%
MBG-20-C_{10}	50 μM	62%	66%

*See Table I.

As shown in Table III, the results obtained indicate that the aminoalkane-terminated oligomers are more effective in producing the desired degree of selective synthesis inhibition than their cognate sequences lacking the terminal alkane. For example, the C_{10} derivative was about 60% more effective than the control unmodified 20 mer in reducing the number of

hemoglobin positive cells. In general, the longer the alkyl chain, the lower the concentration of oligomer required to effect the same % of inhibition.

EXAMPLE 12

The Effect of Fluorescein Terminated Oligonucleotides on the Synthesis of ß-globin Protein in Cultured Cells

Using the methods of synthesis provided in Example 1, 15 to 20 base long oligonucleotides conjugated to a 5'-terminal fluorescein using ethylene diamine as the linker were constructed. This material has the further advantage that uptake of the oligomer into the cells can be monitored by fluorescence microscopy which provides further evidence of the cellular fate of the product. Purified fluorescent oligomers were tested for their effectiveness in preventing the specific expression of hemoglobin in MEL cells induced to produce hemoglobin. The results are shown in Table IV. The protocol for the test is given in Example 10.

TABLE IV

THE EFFECT OF FITC CONJUGATION ON THE INHIBITION OF HEMOGLOBIN SYNTHESIS IN CULTURED CELLS

Oligomer*		% Viable Cells	Inhibition of Benzidine* Cells
DMSO Control		53%	0%
MBG-20 Antisense	50 µM	73%	35%
MBG-20-C_2-FITC	50 µM	68%	45%
MBG-20-C_6-FITC	50 µM	76%	36%
MBG-20-C_{10}-FITC	50 µM	72%	52%

*See Table I.

As shown in Table IV, the results obtained indicate that the fluorescein-terminated oligomers are at

least as effective in producing selective inhibition of hemoglobin synthesis as their cognate control sequences lacking the FITC. Further, fluorescence microscopy of the treated cells showed enhanced fluorescence due to fluoresceinated oligomer uptake. These cells were then isolated, washed several times in physiological saline and lysed by freeze thawing several times in water. The resultant solution was centrifuged to remove cell debris and the amount of fluoresceinated oligomer present quantitated in an Aminco spectrofluorometer. The results obtained showed that the treated cells assimilated an average of 10^7 molecules of fluoresceinated oligomer per cell. This is about 10 times higher than cellular uptake of similar DNA oligomers (i.e lacking the solubility) moiety of about 10^6 molecules per cell.

Thus it can be seen that the addition of a hydrophobic moiety, in this case fluorescein, to the oligomer results in substantially increased cellular uptake of the oligomer without affecting its ability to selectively block protein synthesis.

EXAMPLE 13

The Effect of Polyethylene Glycol Terminated Oligonucleotides on the Synthesis of β-globin Protein in Cultured Cells

Using the methods of synthesis provided in the previous examples, 20 base long oligonucleotides conjugated to a 5'-terminal polyethylene glycol were constructed as described in Example 4. These molecular conjugates were purified and tested for their effectiveness in preventing the specific expression of hemoglobin as described in Example 10.

TABLE V

THE EFFECT OF POLYETHYLENE GLYCOL CONJUGATION ON THE INHIBITION OF HEMOGLOBIN SYNTHESIS IN CULTURED CELLS

Oligomer Conjugate*		Viable Cells (% of Control)	Inhibition of Benzidine* Cells
DMSO Control		33%	0%
MBG-15 Antisense	100 µM	50%	25%
MBG-15-C_2	100 µM	60%	22%
PEG(ss)	100 µM	43%	24%
MBG-20 + PEG(ss)	100 µM	43%	78%
DMSO Control		65%	0%
MBG-20-PEG(ss)	15 µM	0%	95%
	5 µM	62%	52%
	1 µM	nd	-2%
	0.1 µM	64%	-5%

*See Table I.

As shown in Table V, the results obtained show that oligomers conjugated to polyethylene glycol are more effective in producing the desired degree of selective synthesis inhibition than controls. The polyethylene glycol conjugate in this experiment was found to be approximately 10 times more active in preventing the expression of hemoglobin than the control combination of the 20 mer and polyethylene glycol. It is also interesting to note that the simple addition of polyethylene glycol to the medium increases the effectiveness of the added control antisense oligomer, in consonance with the increased effectiveness observed for the PEG conjugates.

It is evident from the above results that the novel conjugates of the subject invention provide substantial advantages in enhancing the efficiency in which transcriptional mechanisms may be modulated. In accordance with the subject invention, a wide variety of cellular, both prokaryotic and eukaryotic, as well

as viral, physiological processes may be regulated. The compositions can be used in vitro and in vivo. In the former, systems can be studied, mammalian cells protected from mycoplasma, phenotypes modified, and the like. In the latter, the compositions can be used for therapy in inhibiting the proliferation of pathogens, selectively inhibiting certain classes of cells, e.g., B-cells and T-cells, or the like.

All publications and patent applications mentioned in this specification are indicative of the level of skill of those skilled in the art to which this invention pertains. All publications and patent applications are herein incorporated by reference to the same extent as if each individual publication or patent application was specifically and individually indicated to be incorporated by reference.

Although the foregoing invention has been described in some detail by way of illustration and example for purposes of clarity of understanding, it will be obvious that certain changes and modifications may be practiced within the scope of the appended claims.

WHAT IS CLAIMED IS:

1. A method for inhibiting the maturation or translation of a messenger RNA in a cell, said method comprising:

contacting said cell with a composition comprising an oligonucleotide sequence complementary to a transcription product of said cell and a group covalently linked to said oligonucleotide sequence to provide an amphiphilic molecule, whereby said composition migrates into the cell interior resulting in the inhibition of maturation and/or translation of said transcription product.

2. A method according to Claim 1, wherein said cell is in culture and said composition is introduced into the nutrient medium.

3. A method according to Claim 1, wherein said oligonucleotide is of from about 6 to 30 nucleotides.

4. A method according to Claim 3, wherein at least one of said oligonucleotides has a phosphate as the phosphorus moiety.

5. A method according to Claim 3, wherein at least one of said oligonucleotides has a phosphonate with an alkyl group of from 1 to 3 carbon atoms as the phosphorus moiety.

6. A method according to Claim 1, wherein said group is a hybridphobic aromatic group.

7. A method according to Claim 7, wherein said aromatic group is a trityl group.

8. A method according to Claim 7, wherein said aromatic group is a fluorescein group.

9. A method according to Claim 1, wherein said group is a polyalkyleneoxy group, wherein said alkylenes are of from 2 to 10 carbon atoms.

10. A method according to Claim 9, wherein said polyalkyleneoxy group is from about 6 to 200 units.

11. A cell comprising a composition comprising an oligonucleotide sequence complementary to a transcription product of said cell and an amphiphilic or hydorphobic group covalently linked to said oligonucleotide sequence to provide an amphiphilic molecule.

12. A cell according to Claim 11, wherein said cell is in culture.

13. A composition of matter comprising:

an oligonucleotide sequence of at least six nucleotides complementary to a transcriptional product of a cell;

an amphiphilic group comprising a polyalkyleneoxy group, wherein said alkylenes are of from 2 to 10 carbon atoms;

a linker of at least one atom covalently bonded to said oligonucleotide sequence and to said amphiphilic group.

14. A composition of matter according to Claim 13, wherein said oligonucleotide is of from about 6 to 30 nucleotides.

15. A composition of matter according to Claim 13, wherein at least one of said oligonucleotides has a phosphate as the phosphorus moiety.

16. A composition of matter according to Claim 13, wherein at least one of said oligonucleotides has a phosphonate with an alkyl group of from 1 to 3 carbon atoms as the phosphorus moiety.

17. A composition of matter according to Claim 13, wherein said linking group includes at least one of an amino, quinone, thioether, or amide group.

18. A composition of matter according to Claim 13, wherein said oligonucleotide sequence is complementary at least in part to a non-coding region.

19. A composition of matter according to Claim 13, wherein said oligonucleotide sequence is complementary at least in part to a coding region.

INTERNATIONAL SEARCH REPORT

International Application No. PCT/US88/02009

I. CLASSIFICATION OF SUBJECT MATTER (if several classification symbols apply, indicate all) [6]

According to International Patent Classification (IPC) or to both National Classification and IPC

IPC(4): C12N 5/00; C12N 5/02, C12P 19/34
U.S. 435/243, 435/91

II. FIELDS SEARCHED

Minimum Documentation Searched [7]

Classification System	Classification Symbols
US	435/6,91,243 530/358 536/27, 28, 29

Documentation Searched other than Minimum Documentation to the Extent that such Documents are Included in the Fields Searched [8]

Chemical Abstracts Data Base (CAS) 1967-1988
Keywords: amphiphilic, nucleic acid

III. DOCUMENTS CONSIDERED TO BE RELEVANT [9]

Category *	Citation of Document, [11] with indication, where appropriate, of the relevant passages [12]	Relevant to Claim No. [13]
E,Y	U.S, A, 4,757,141, 12 July 1988, (FUNG ET AL.), see column 1, lines 8-31, and column 2, lines 40-63.	1-19
Y	U.S., A, 4,511,713, 16 April 1985, (MILLER ET AL.), see column 2, lines 13-68, and column 3, lines 1-48.	1-19
Y	U.S., A, 4,587,044 6 May 1986, (MILLER ET AL.), see column 1, line 25 - column 3, line 60.	1-19
Y	FR 2,556,726 21 June 1985 (CALIFORNIA INSTITUTE OF TECHNOLOGY). page 3, line 23 - page 5, line 8.	1-19
Y	GB, 2,153,256 A, 21 August 1985, (CALIFORNIA INSTITUTE OF TECHNOLOGY), see page 1, line 125 - page 2, line 47.	1-19

* Special categories of cited documents: [10]

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

IV. CERTIFICATION

Date of the Actual Completion of the International Search	Date of Mailing of this International Search Report
12 September 1988	08 NOV 1988
International Searching Authority	**Signature of Authorized Officer**
ISA/US	STEPHANIE SEIDMAN, Ph.D., J.D.

III. DOCUMENTS CONSIDERED TO BE RELEVANT (CONTINUED FROM THE SECOND SHEET)

Category *	Citation of Document, with indication, where appropriate, of the relevant passages	Relevant to Claim No
Y	NUCLEIC ACIDS RESEARCH, (Oxford, England), Volume 13, number 7, issued 11 April 1985, (SMITH ET AL.), "Synthesis of oligonucleotides containing an aliphatic amino group at the 5' terminus: synthesis of fluorescent DNA primers for use in DNA sequence analysis", see abstract.	1-19
Y	BIOCHEMISTRY, (Easton, Pennsylvania, U.S.A.), Volume 24, number 22, issued 22 October 1985, (BLAKE ET AL.), "Inhibition of Rabbit Globin mRNA Translation by Sequence-Specific Oligodeoxyribonucleotides", see pages 6132-6133.	1-19
Y	BIOCHEMISTRY, (Easton, Pennsylvania, U.S.A.), Volume 13, number 24, issued 19 November 1974, (BARRETT ET AL.), "Inhibitory Effect of Complex Formation with Oligodeoxyribonucleotide Ethyl Phosphotriesters on Transfer Ribonucleic Acid Aminoacylation", See page 4897.	1 -19
Y	NUCLEIC ACIDS RESEARCH, (Oxford, England), Volume 13, number 5, issued March 1985, (CHOLLET ET AL.), "Biotin-labeled synthetic oligo-deoxlyribonucleotides: chemical synthesis and use as hybridization probes," see page 1529.	1-19
Y	NUCLEIC ACIDS RESEARCH, (Oxford, England), Volume 13, number 12, issued June 1985, (CONNOLLY,) "Chemical synthesis of oligonucleotides containing a free sulphydryl group and subsequent attachment to thiol specific probes", see page 4485.	1-19
Y	NUCLEIC ACIDS RESEARCH, (Oxford, England), Volume 11, number 19, issued September 1983, (CHU ET AL.) "Derivatization of unprotected polynucleotides", see page 6513.	1-19
Y	SCIENCE, (Washington, D.C., U.S.A,) Volume 230, issued 18 October 1985, (CARUTHERS), "Gene Synthesis Machines DNA Chemistry and its Uses", see page 281.	1-19

Identification of the Cystic Fibrosis Gene: Cloning and Characterization of Complementary DNA

JOHN R. RIORDAN, JOHANNA M. ROMMENS, BAT-SHEVA KEREM, NOA ALON, RICHARD ROZMAHEL, ZBYSZKO GRZELCZAK, JULIAN ZIELENSKI, SI LOK, NATASA PLAVSIC, JIA-LING CHOU, MITCHELL L. DRUMM, MICHAEL C. IANNUZZI, FRANCIS S. COLLINS, LAP-CHEE TSUI

Overlapping complementary DNA clones were isolated from epithelial cell libraries with a genomic DNA segment containing a portion of the putative cystic fibrosis (CF) locus, which is on chromosome 7. Transcripts, approximately 6500 nucleotides in size, were detectable in the tissues affected in patients with CF. The predicted protein consists of two similar motifs, each with (i) a domain having properties consistent with membrane association and (ii) a domain believed to be involved in ATP (adenosine triphosphate) binding. A deletion of three base pairs that results in the omission of a phenylalanine residue at the center of the first predicted nucleotide-binding domain was detected in CF patients.

CYSTIC FIBROSIS (CF) IS AN AUTOSOMAL RECESSIVE GENETIC disorder affecting a number of organs, including the lung airways, pancreas, and sweat glands (*1*). Abnormally high electrical potential differences have been detected across the epithelial surfaces of the CF respiratory tract, including the trachea and nasal polyps, as well as across the walls of CF sweat gland secretory coils and reabsorptive ducts (*2*). The basic defect has been associated with decreased chloride ion conductance across the apical membrane of the epithelial cells (*3*). That the defect also appeared to persist in cultured cells derived from several epithelial tissues suggested that the CF gene is expressed in these cells (*4*). More recently, patch clamp studies showed that this defect is probably due to a failure of an outwardly rectifying anion channel to respond to phosphorylation by cyclic AMP–dependent protein kinase (protein kinase A) or protein kinase C (*5*). Although progress has been made in the isolation of polypeptide components of an epithelial chloride channel that mediates conductance (*6*), their relation to the kinase-activated pathway and CF has yet to be established, and the basic biochemical defect in CF remains unknown.

Molecular cloning experiments have permitted the isolation of a large, contiguous segment of DNA spanning at least four transcribed sequences from a region thought to contain the CF locus (*7*). These sequences were initially identified on the basis of their ability to detect conserved sequences in other animal species by DNA hybridization and were subsequently characterized by RNA hybridization experiments, cDNA isolation, and direct DNA sequence analysis (*7*). Three of the transcribed regions were excluded from being the CF locus by earlier genetic or DNA sequence analyses (*7, 8*). The fourth one, as shown by genetic analysis (*9*) and DNA sequencing analysis presented below, corresponds to a portion of the CF gene locus.

Isolation of cDNA clones. Two DNA segments (E4.3 and H1.6) that detected cross-species hybridization signals (*7*) were used as probes to screen cDNA libraries made from several tissues and cell types (*10*). After screening seven different libraries, one single clone (10-1) was isolated with H1.6 from a cDNA library made from the cultured epithelial cells of the sweat glands of an unaffected (non-CF) individual (*10*).

DNA sequencing showed that 10-1 contained an insert of 920 base pairs (bp) in size and one potential, long open reading frame (ORF). Since one end of the sequence shared perfect sequence identity with H1.6, it was concluded that the cDNA clone was probably derived from this region. The DNA sequence in common was, however, only 113 bp long (Figs. 1 and 2). This sequence in fact corresponded to the first axon of the putative CF gene. The short sequence overlap thus explained the weak hybridization signals in library screening and our inability to detect transcripts in RNA gel-blot analysis. In addition, the orientation of the transcription unit was tentatively established on the basis of alignment of the genomic DNA sequence with the presumptive ORF of 10-1.

Since the corresponding transcript was estimated to be about 6500 nucleotides in length by RNA gel-blot hybridization experiments, further cDNA library screening was required in order to clone the remainder of the coding region. As a result of several successive screenings with cDNA libraries generated from the colon carcinoma cell line T84, normal and CF sweat gland cells, pancreas,

J. R. Riordan, N. Alon, and Z. Grzelczak are in the Department of Biochemistry, Research Institute, The Hospital for Sick Children, Toronto, Ontario M5G 1X8, Canada. J. M. Rommens, B. Kerem, R. Rozmahel, J. Zielenski, S. Lok, N. Plavsic, J.-L. Chou, and L.-C. Tsui are at the Department of Genetics, Research Institute, The Hospital for Sick Children, Toronto, Ontario M5G 1X8, Canada. M. L. Drumm, M. C. Iannuzzi, and F. S. Collins are at the Howard Hughes Medical Institute and Departments of Internal Medicine and Human Genetics, University of Michigan, Ann Arbor, MI 48109. J. R. Riordan is also a member of the Departments of Biochemistry and Clinical Biochemistry, and L.-C. Tsui is also a member of the Departments of Medical Genetics and Medical Biophysics, University of Toronto, Toronto, Ontario M5S 1A8, Canada.

and adult lungs, 18 additional clones were isolated (Fig. 1). DNA sequence analysis revealed that none of these cDNA clones corresponded to the length of the observed transcript, but it was possible to derive a consensus sequence based on overlapping regions. Further cDNA clones corresponding to the 5′ and 3′ ends of the transcript were derived from 5′ and 3′ primer-extension experiments (Fig. 1). Together, these clones span about 6.1 kb and contain an ORF capable of encoding a protein of 1480 amino acids (Fig. 2).

It was unusual that most of the cDNA clones isolated here contained sequence insertions at various locations (Fig. 1). While many of these extra sequences corresponded to intron regions reverse-transcribed during the construction of the cDNA, as revealed on alignment with genomic DNA sequences, the identities of several others were uncertain because they did not align with sequences at the corresponding exon-intron junctions, namely, the sequences at the 5′ ends of clones 13a and T16-1 and at the 5′ and 3′ ends of T11, and the insertions between exons 3 and 4 in 13a and between exons 10 and 11 in T16-4.5 (legend to Fig. 1). More puzzling were the sequences corresponding to the reverse complement of exon 6 at the 5′ end of 11a and the insertion of a segment of a bacterial transposon in clone C16-1; none of these could be explained by mRNA processing errors.

In that the number of recombinant cDNA clones for the putative CF gene detected in the library screening was much less than would have been expected from the abundance of transcripts estimated from RNA hybridization experiments, it seemed probable that the clones that contained aberrant structures were preferentially retained while the proper clones were lost during propagation. Consistent with this interpretation, poor growth was observed for most of our recombinant clones isolated, regardless of the vector used.

RNA analysis. To visualize the transcript of the putative CF gene, we used RNA gel-blot hybridization with the 10-1 cDNA as the probe (Fig. 3). The analysis revealed a prominent band, about 6.5 kb in size, in T84 cells. Identical results were obtained with other cDNA clones as probes. Similar, strong hybridization signals were also detected in pancreas and primary cultures of cells from nasal polyps, suggesting that the mature mRNA of the putative CF gene is about 6.5 kb. Minor hybridization signals, probably representing degradation products, were detected at the lower size ranges, but they varied between different experiments. On the basis of the hybridization band intensity and comparison with those detected for other transcripts under identical experimental conditions, it was estimated that the putative CF gene transcripts constituted about 0.01 percent of total mRNA in T84 cells.

Additional tissues were analyzed by RNA gel-blot hybridization in an attempt to correlate the expression pattern of the putative CF gene and the pathology of CF. Transcripts, all of identical size, were found in lung, colon, sweat glands (cultured epithelial cells), placenta, liver, and parotid gland, but the signal in these tissues was generally weaker than that detected in the pancreas and nasal polyps (Fig. 3). Intensity varied among different preparations; for example, hybridization in kidney was not detectable in the preparation shown in Fig. 3 but was clearly discernible subsequently. Transcripts were not detected in the brain or adrenal gland, nor in skin fibroblast and lymphoblast cell lines.

Thus, expression of the putative CF gene appeared to occur in many of the tissues examined, with higher levels in those tissues severely affected in CF. While this epithelial tissue–specific expression pattern is in good agreement with the disease pathology, no significant difference was detected in the amount or size of transcripts from CF and control tissues (Fig. 3), consistent with the assumption that CF mutations are subtle changes at the nucleotide level.

Characterization of cDNA clones. As indicated above, a contig-

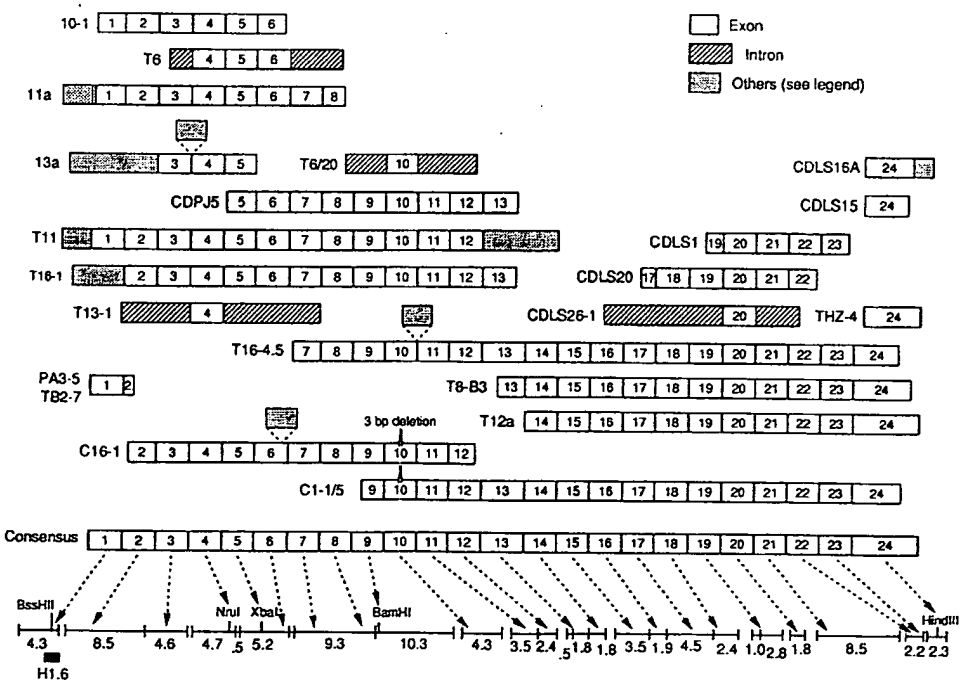
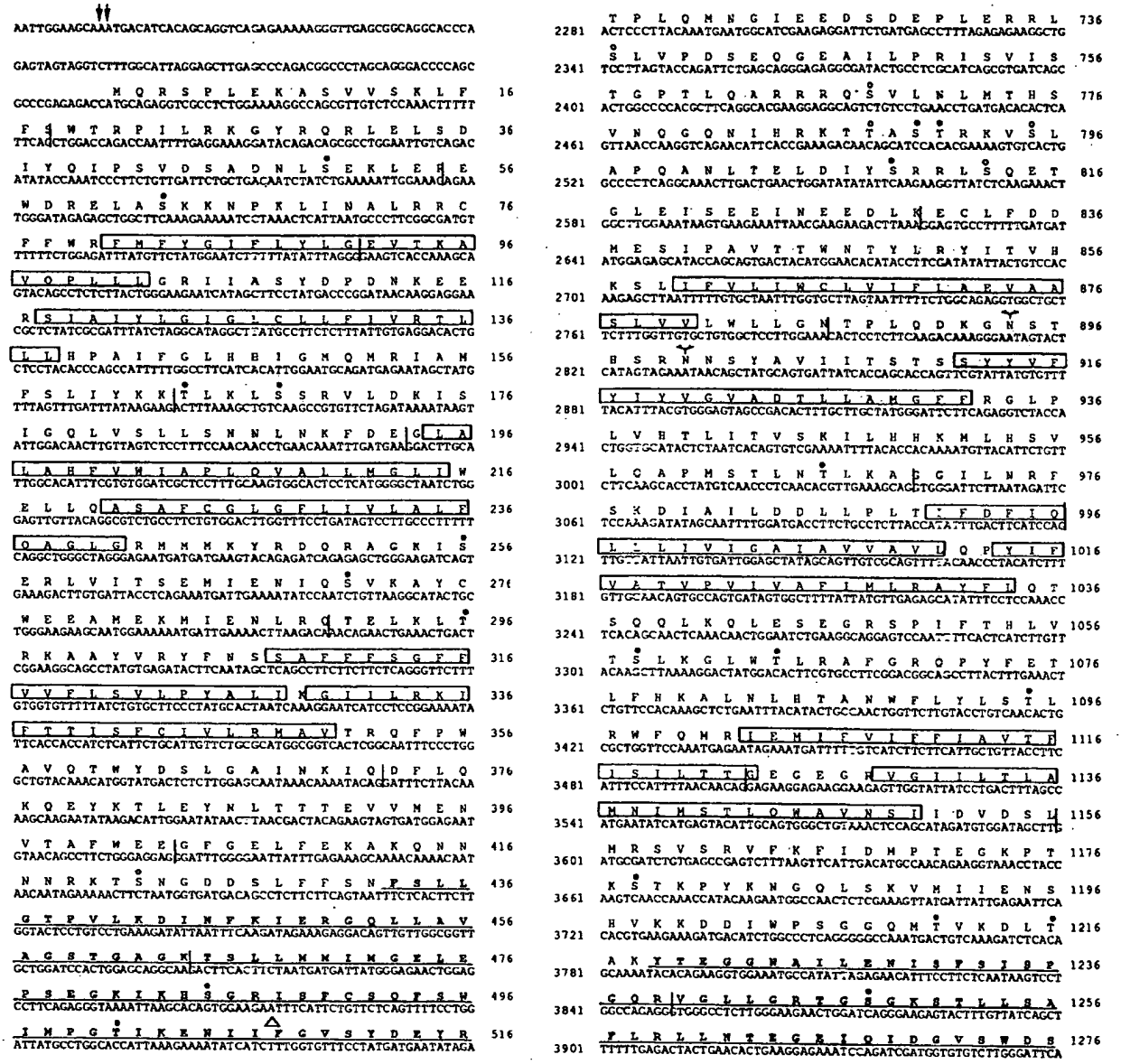


Fig. 1. Overlapping cDNA clones aligned with genomic DNA fragments. The cDNA clones are represented by open boxes with exons indicated. The corresponding genomic Eco RI fragments are schematically presented on the bottom, with lengths in kilobases. The hatched boxes denote intron sequences, and stippled boxes represent other sequences as outlined below. The filled box in the lower left is the position of the clone H1.6, which was used to isolate the first cDNA clone 10-1 from a normal (N) sweat gland library (*10*). The definitive restriction sites used for the alignment of cDNA and genomic fragments are indicated. Clones T6, T6/20, T11, T16-1, T13-1, T16-4.5, T8-B3, and T12a were isolated sequentially from the T84 cell library (*10*). Clones isolated from the human lung cDNA library (*10*) are designated with the prefix CDL. CDPJ5 is derived from a pancreas library (*10*). The CF sweat gland cDNA clones, C16-1 and C1-1/5, together cover all but exon 1 and a portion of the 3′ untranslated region. Both clones revealed a 3-bp deletion in exon 10. Clones that contain intron sequences are CDLS26-1, T6/20, and T13-1. Clones T11, T16-4.5, CDLS16A, 11a, and 13a contain extraneous sequences of unknown origin at positions indicated. Clone C16-1 contains a short insertion corresponding to a portion of the γ transposon of *E. coli*. Both PA3-5 and TB2-7 are 5′ extension clones generated from pancreas and T84 RNA by the anchored PCR technique (*12*), respectively. THZ-4 is a 3′ extension clone obtained from T84 RNA. Both T12a and THZ-4 contain a polyadenylation signal and a poly(A)$^+$ tail.



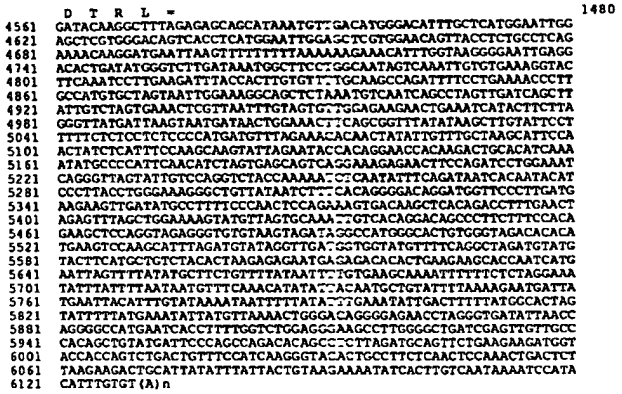


Fig. 2. Nucleotide sequence of cDNA encoding the CF transmembrane conductance regulator together with the deduced amino acid sequence. DNA sequencing was performed by the dideoxy chain termination method (*34*) with ^{35}S-labeled nucleotides or by the Dupont Genesis2000 automatic DNA sequencer. Numbers on the left of columns indicate base positions and numbers on the right amino acid residue positions. The first base position corresponds to the first nucleotide in the 5′ extension clone PA3-5, which is one nucleotide longer than TB2-7 (*12*). The 3′ end and the noncoding sequence are shown above [nucleotides 4561 to 6129 plus the poly(A)$^+$ tail]. Arrows indicate position of transcription initiation site by primer extention analysis (*11*). Nucleotide 6129 is followed by a poly(A) tract. Positions of exon junctions are indicated by vertical lines. Potential membrane-spanning segments ascertained with the use of the algorithm of Eisenberg *et al.* (*35*) are enclosed in boxes. Amino acids comprising putative ATP-binding folds are underlined. Possible sites of phosphorylation (*21*) by protein kinases A or C are indicated by open and closed circles, respectively. The open triangle indicates the position at which 3 bp are deleted in CF. Abbreviations for the amino acid residues are: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.

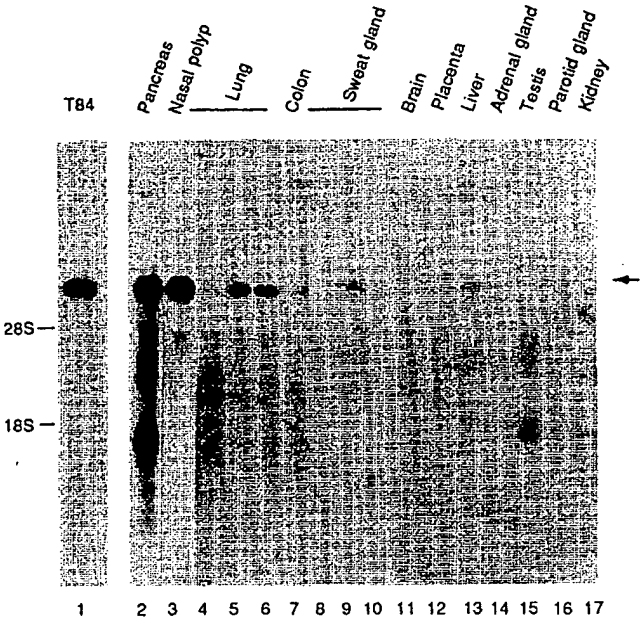


Fig. 3. RNA gel-blot analysis. Hybridization by the cDNA clone 10-1 to a 6.5-kb transcript is shown in the tissues indicated. RNA samples were prepared from cells and tissue samples obtained from surgical pathology or at autopsy according to the methods described in (*10*). Total RNA (10 μg) from each tissue and 1 μg of poly(A)$^+$ RNA from T84 cells were separated on formaldehyde gels and transferred onto nylon membranes (Zetaprobe, Bio-Rad), which were hybridized with DNA probes labeled to high specific activity by the random priming method (*36, 37*). The positions of the 28*S* and 18*S* rRNA bands are indicated.

uous coding region of the CF locus could be deduced from overlapping cDNA clones. Since most of the cDNA clones were apparently derived from unprocessed transcripts, further studies were performed to ensure the authenticity of the consensus sequence. Each cDNA clone was first tested for chromosome localization by hybridization analysis with a human-hamster somatic cell hybrid containing a single human chromosome 7 and by pulsed field gel electrophoresis (*7*). The ones that did not map to the correct region on chromosome 7 were not pursued. Fine restriction enzyme mapping was then performed for each clone. While overlapping regions were clearly identified for most of the clones, many contained single copy, additional regions not readily recognizable by restriction enzyme analysis.

The cDNA was further characterized in gel hybridization experiments with genomic DNA. Five to six different restriction fragments could be detected with the 10-1 cDNA in Eco RI– or Hind III–digested total human DNA and a similar number of fragments with several other cDNA clones, suggesting the presence of multiple exons for the putative CF gene. The hybridization studies also identified the cDNA clones with unprocessed intron sequences when they showed preferential hybridization to a smaller subset of genomic DNA fragments with relatively greater intensities. For the confirmed cDNA clones, their corresponding genomic DNA segments were isolated (*7*) and the exons and exon-intron boundaries were sequenced. In all, 24 exons were identified (Fig. 2). Physical mapping experiments (*7*) showed that the gene locus spanned about 250 kb.

The 5′ terminus of the transcript was determined by primer extension (*11*). A modified polymerase chain reaction, anchored PCR (*12*), was also used to facilitate cloning of the 5′ end sequences.

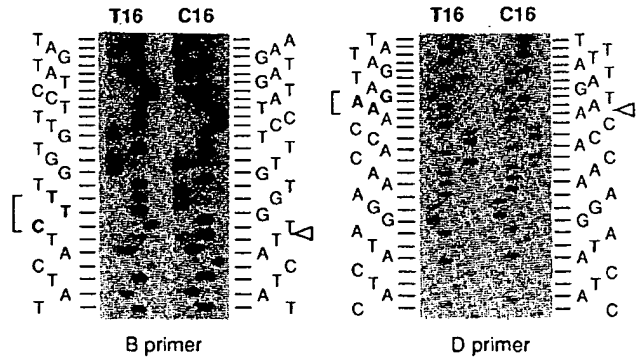


Fig. 4. DNA sequence around the ΔF_{508} deletion. The normal sequence from base position 1627 to 1651 (from cDNA T16-1) is shown beside the CF sequence (from cDNA C16-1). The left panel shows the sequences from the coding strands obtained with the B primer (5′-GTTTTCCTGGATTATGCCTGGGCAC-3′) and the right panel those from the opposite strand with the D primer (5′-GTTGGCATGCTTTGATGACGCTTC-3′). The brackets indicate the three nucleotides in the normal that are absent in CF (arrowheads). Sequencing was performed as described in (*34*).

Two independent 5′ extension clones, one from pancreas and the other from T84 RNA, were characterized by DNA sequencing and differed by only 1 base in length, thus establishing the most probable initiation site for the transcript (Fig. 2). Since the initial cDNA clones did not contain a poly(A)$^+$ tail indicative of the end of a mRNA, anchored PCR was also applied to the 3′ end of the transcript (*12*). The results derived from the use of several different 3′-extending oligonucleotides were consistent with the interpreta-

tion that the end of the transcript was about 1.2 kb downstream of the Hind III site at nucleotide position 5027 (Fig. 2).

The complete cDNA sequence spans 6129 base pairs excluding the poly(A)$^+$ tail at the end of the 3′ untranslated region and it contains an ORF capable of encoding a polypeptide of 1480 amino acids (Fig. 2). An ATG (AUG) triplet is present at the beginning of this ORF (base position 133–135). Since the nucleotide sequence surrounding this codon (5′-AGACC**AUG**CA-3′) has the proposed features of the consensus sequence (CC)$^{A}_{G}$CC**AUG**G(G) of a eukaryotic translation initiation site (*13*), with a highly conserved A at the −3 position, it is highly probable that this AUG corresponds to the first methionine codon for the putative polypeptide.

Detection of mutation. A comparison between the cDNA sequences derived from CF and unaffected (N) individuals was next conducted. Two clones, C16-1 and C1-1/5, were derived from a CF sweat gland cDNA library and together they spanned almost the entire coding region. The most striking difference between CF and N sequences was a 3-bp deletion (Fig. 4), which would result in a loss of a phenylalanine residue (position 508) in the predicted CF polypeptide. This deletion (ΔF_{508}) was detected in both CF clones. To exclude the possibility that this difference was due to a cloning artifact, sequence-specific oligonucleotides were used to screen DNA samples from CF families. Specific hybridization could be observed for each oligonucleotide probe with genomic DNA amplified by PCR, confirming the presence of corresponding genomic DNA sequences (*9*). Furthermore, the oligonucleotide specific for the 3-bp deletion hybridized to 68 percent of chromosomes carrying a CF mutation but not to any of the normal chromosomes (0/198), an indication that a silent sequence polymorphism was unlikely. Sequence differences found elsewhere among the different cDNA clones probably represented sequence polymorphisms or cDNA cloning artifacts (*14*).

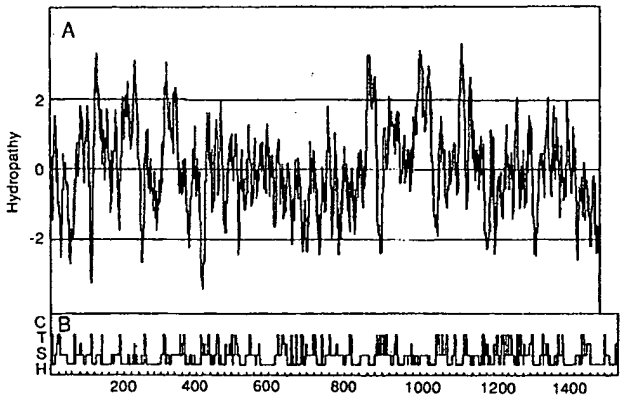


Fig. 5. Hydropathy profile and predicted secondary structures of the CFTR. (**A**) The mean hydropathy index determined according to Kyte and Doolittle (*19*) of nine-residue peptides is plotted against the amino acid number. (**B**) The corresponding positions of features of secondary structure predicted according to Garnier *et al.* (*19*). C, coil; T, turn; S, sheet; H, helix.

Predicted protein structure. Analysis of the sequence of the overlapping cDNA clones (Fig. 2) predicted a polypeptide of 1480 amino acids with a molecular mass of 168,138 daltons. The most characteristic feature of the predicted protein is the presence of two repeated motifs, each of which consists of a domain capable of spanning the membrane several times and sequences resembling consensus nucleotide (ATP)-binding folds (NBF's) (*15*) (Figs. 5 and 6). These characteristics are remarkably similar to those of the mammalian multidrug resistance P-glycoprotein (*16*) and a number of other membrane-associated proteins (as discussed below), suggesting that the predicted CF gene product is likely to be involved in the transport of substances (ions) across the membrane and is probably a member of a membrane protein superfamily (*17*). For the convenience of future discussion and to avoid confusion with the previously named CF protein and CF factor (*18*), we will call the

```
CFTR (N)   FSLLGTPVLKDINFKIERGQLLAVAGSTGAGKTSLLMMIMG  ISFCSQFSWIMPGTIK-ENIIFGVSYD  GEGGITLSGGQRARISLARAVYKDADLYLLDSPFGYLDVLTEK
CFTR (C)   YTEGGNAILENISFSISPGQRVGLLGRTGSGKSTLLSAFLR  DSITLQQWRKAFGVIPQKVFIFSGTFR  VDGGCVLSHGHKQLMCLARSVLSKAKILLLDEPSAHLDPVTYQ
hmdr1 (N)  PSRKEVKILKGLNLKVQSGQTVALVGNSGCGKSTTVQLMQR  IGVVSQEPVLFATTI-AENIRYGRENV  GERGAQLSGGQKQRIAIARALVRNPKILLLDEATSALDTESEA
hmdr1 (C)  PTRPDIPVLQGLSLEVKKGQTLALVGSSGCGKSTVVQLLER  LGIVSQEPILFDCSI-AENIAYGDNSR  GDKGTLLSGGQKQRIAIARALVRQPHILLLDEATSALDTESEK
mmdr1 (N)  PSRSEVQILKGLNLKVKSGQTVALVGNSGCGKSTTVQLMQR  IGVVSQEPVLFATTI-AENIRYGREDV  GERGAQLSGGQKQRIAIARALVRNPKILLLDEATSALDTESEA
mmdr1 (C)  PTRPNIPVLQGLSLEVKKGQTLALVGSSGCGKSTVVQLLER  LGEVSQEPILFDCSI-AENIAYGDNSR  GDKGTQLSGGQKQRIAIARALVRQPHILLLDEATSALDTESEK
mmdr2 (N)  PSRANIKILKGLNLKVKSGQTVALVGNSGCGKSTTVQLLQR  IGVVSQEPVLSFTTI-AENIRYGRGNV  GDRGAQLSGGQKQRIAIARALVRNPKILLLDEATSALDTESEA
mmdr2 (C)  PTRANVPVLQGLSLEVKKGQTLALVGSSGCGKSTVVQLLER  LGIVSQEPILFDCSI-AENIAYGDNSR  GDKGTQLSGGQKQRIAIARALIRQPRVLLLDEATSALDTESEK
pfmdr (N)  DTRKDVEIYKDLSFTLLKEGKTYAFVGESGCGKSTILKLIE  IGVVSQDPLLFSNSI-KNNIKYSLYSL  GSNASKLSGGQKQRISIARAIMRNPKILILDEATSSLDNKSEY
pfmdr (C)  ISRPNVPIYKNLSFTCDSKKTTAIVGETGSGKSTFMNLLLR  FSIVSQEPMLFNMSI-YENIKFGREDA  PYGKS-LSGGQKQRIAIARALLREPKILLLDEATSSLDSNSEK
STE6 (N)   PSRPSEAVLKNVSLNFSAGQFTFIVGKSGSGKSTLSNLLLR  ITVVEQRCTLFNDTL-RKNILLGSTDS  GTGGVTLSGGQQQRVAIARAFIRDTPILFLDEAVSALDIVRRN
STE6 (C)   PSAPTAFVYKNMNFDMFCGQTLGIIGESGTGKSTLVLLLTK  ISVVEQKPLLFNGTI-RDNLTYGLQDE  RIDTTLLSGGQAQRLCIARALLRKSKILILDECTSALDSVSSS
hlyB       YKPDSPVILDNINISIKQGEVIGIVGRSGSGKSTLIKLIQR  VGVVLQDNVLLNRSI-IDNISLAPGMS  GEQGAGLSGGQRQRIAIARALVNNPKILIFDEATSALDYASEH
White      IPAPRKHLLKNVCGVAYPGELLAVMGSSGAGKTTLLNALAF  RCAYVQQDDLFIGLIARENLIFQAMVR  PGRVKGLSGGERKRLAFASEALTDPPLLICDEPTSGLDSFTAH
MbpX       KSLGNLKILDRVSLYVPKFSLIALLGPSGSGKSSLLRILAG  MSFVFQHYALFKHMTVYENISFGLRLR  FEYPAQLSGGQKQRVALARSLAIQPDLLL-DEPFGALDGELRR
BtuD       QDVAESTRLGPLSGEVRAGRILHLVGPNGAGKSTLLARIAG  YLSQQQTPPFATPVWHYLTLHQHDKTR  GRSTNQLSGGEWQRVRLAAVVLQITLLLLDEPMNSLDVAQQSA
PstB       FYYGKFHALKNINLDTAKNQVTAFIGPSGCGKSTLLRTFNK  VGMVFQKPTPFPMSI-YDNIAFGVRLF  HQSGYSLSGGQQQRLCIARGIAIRPEVLLLDEPCSALDPISTG
hisP       RRYGGHEVLKGVSLQARAGDVISIIGSSGSGKSTFLRCINF  GIMVFQHFNLWSHMTVLENVMEAPIQV  GKYPVHLSGGQQQRVSIARALAMEPDVLLFDEPTSALDPELVG
malK       KAWGEVVVSKDINIDIHEGEFVVFVGPSGCGKSTLLRMIAG  VGMVFQSYALYPHLSVAENMSFGLKPA  DRKPKALSGGQRQRVAIGRTLVAEPSVFLLDEPLSNLDAALRV
oppD       TPDGDVTAVNDLNFTLRAGETLGIVGESGSGKSQTAFALMG  ISMIFQDPMTSLNPYMRVGEQLMEVLM  KMYPHEFSGGMRQRVMIAMALLCRPKLLIADEPTTALDVTVQA
oppF       QPPKTLKAVDGVTLRLYEGETLGVVGESGCGKSTFARAIIG  IQMIFQDPLASLNPRMTIGEIIAEPLR  NRYPHEFSGGQCQRIGIARALILEPKLIICDDAVSALDVSIQA
RbsA (N)   KAVPGVKALSGAALNVYPGRVMALVGENGAGKSTMMKVLTG  AGIIHQELNLIPQLTIAENIFLGREFV  DKLVGDLSIGDQQMVEIAKVLSFESKVIIMDEPTCALIDTETE
RbsA (C)   VDNLCGPGVNDVSFTLRRGEILGVSGLMGAGRTELMKVLYG  ISEDRKRDGLVLGMSVKENMSLTALRY  EQAIGLLSGGNQQKVAIARGLMTRPKVLILDEPTPGVDVGAKK
UvrA       LTGARGNNLKDVTLTLPVGLFTCITGVSGSGKSTLINDTLF  TYTGVFTPVRELFAGVPESRARGYTPG  GQSATTLSGGEAQRVKLARELSKRGLYILDEPTTGLHFADIQQ
NodI       KSYGGKIVVNDLSFTIAAGECFGLLGPNGAGKSTIIRMILG  IGIVSQEDNLDLEFTVRENLLVYGRYF  NTRVADLSGGMKRRLTLAGALINDPQLLILDEPTTGLDPHARH
FtsE       AYLGGRQALQGVTFHMQPGEMAFLTGHSGAGKSTLLKLICG  IGMIFQDHRLLMDRTVYDNVAIPLIA  KNFPIQLSGGEQQRVGIARAVVNKPAVLLADEPTGNLDDALSE
```

Fig. 6. Alignment of the three most conserved segments of the amino acid sequences (single letter code) of the extended NBF's of CFTR with comparable regions of other proteins. These three segments consist of residues 433 to 473, 488 to 513, and 542 to 584 of the amino-terminal (N) half and 1219 to 1259, 1277 to 1302, and 1340 to 1382 of the carboxyl-terminal (C) half of CFTR. The heavy overlining points out the regions of greatest similarity. The star indicates the position corresponding to the phenylalanine that is deleted in CF. Additional general homology can be seen even with the introduction of very few gaps. The other sequences are of proteins involved in multidrug resistance in human (hmdr1), mouse (mmdr 1 and 2) (*16*), and *Plasmodium falciparum* (pfmdr) (*38*); the α-factor pheromone export system of yeast (STE6) (*39*); the hemolysin (hlyB) system of *E. coli* (*22*); screening of eye pigments in *Drosophila* (White) (*23*); an unknown liverwort chloroplast function (Mbpx) (*25*); vitamin B12 transport in *E. coli* (BtuD) (*24*); phosphate transport in *E. coli* (PstB) (*40*); histidine transport in *Salmonella typhimurium* (hisP) (*41*); maltose transport in *E. coli* (malK) (*42*); oligopeptide transport in *S. typhimurium* (oppD and oppF) (*43*); ribose transport in *E. coli* (RbsA) (*44*). UvrA is one component of an *E. coli* DNA repair system (*45*); NodI is a gene product involved in nodulation in *Rhizobium* (*46*); FtsE is a protein that contributes to the regulation of cell division (*47*). In addition to these proteins that contain this long NBF, there is a large number of others that contain the two short nucleotide binding motifs A and B initially pointed out by Walker *et al.* (*48*). Further, there are other proteins containing only motif A or B (*49*).

putative CF gene product the cystic fibrosis transmembrane conductance regulator (CFTR).

Each of the predicted membrane-associated regions of CFTR consists of six hydrophobic segments capable of spanning a lipid bilayer (*19*), which are followed by a large hydrophilic region containing the NBF's (Fig. 5). On the basis of sequence alignment with other nucleotide-binding proteins, each of the putative NBF's in CFTR comprises at least 150 residues (Fig. 6). The single residue deletion (ΔF_{508}) detected in most of the CF patients is in the first NBF, between the two most highly conserved segments within this sequence. The amino acid sequence identity between the region surrounding the ΔF_{508} mutation and the corresponding regions of several other proteins suggests that this region is of functional importance (Fig. 6). A hydrophobic amino acid, usually one with an aromatic side chain, is present in most of these proteins at the position corresponding to Phe^{508} of CFTR.

Despite the overall symmetry in the two-motif structure of the protein and the sequence conservation of the NBF's, sequence identity between the two motifs of the predicted CFTR protein is modest. The strongest identity is between sequences at the carboxyl ends of the NBF's. Of the 66 residues aligned within these regions, 27 percent are identical and 11 percent are functionally similar. The overall, weak internal sequence identity is in contrast to the much higher degree (>70 percent) in P-glycoprotein for which a sequence duplication hypothesis has been proposed (*16*). The lack of conservation in the relative positions of the exon-intron boundaries in the CF gene also argues against recent exon duplication as a mechanism in the evolution of this gene (Fig. 2).

Since there is apparently no signal-peptide sequence at the amino terminus of CFTR (Fig. 7), the highly charged hydrophilic segment preceding the first transmembrane sequence is probably oriented in the cytoplasm. Each of the two sets of hydrophobic helices are expected to form three traversing loops across the membrane and little of the sequence of the entire protein is expected to be exposed to the exterior surface, except the region between transmembrane segments 7 and 8. It is of interest that the latter region contains two potential sites for N-linked glycosylation (*20*).

A highly charged cytoplasmic domain can be identified in the middle of the predicted CFTR polypeptide, linking the two halves of the protein. This domain, named the R domain, is operationally defined by a single large exon in which 69 of the 241 amino acids are polar residues arranged in alternating clusters of positive and negative charges. Moreover, nine of the ten sites at which there are consensus sequences for phosphorylation by protein kinase A and seven of the potential substrate sites for protein kinase C found in CFTR are located in this exon (*21*).

Properties of CFTR could be further derived from comparison to other membrane-associated proteins (Fig. 6). In addition to the overall structural similarity with P-glycoproteins, each of the two predicted motifs in CFTR shows resemblance to the single motif structure of hemolysin B of *Escherichia coli* (*22*) and the product of the White gene of *Drosophila* (*23*). These proteins are involved in the transport of the lytic peptide of the hemolysin system and of eye pigment molecules, respectively. The vitamin B12 transport system of *E. coli*, BtuD (*24*), and MbpX (*25*), which is a liverwort chloroplast gene product whose function is unknown, also have a similar structural motif. Further, CFTR shares structural similarity with several of the periplasmic solute transport systems of Gram-negative bacteria, where the transmembrane region and the ATP-binding folds are contained in separate proteins that function in concert with a third substrate-binding polypeptide (*26*).

The overall structural arrangement of the transmembrane domains in CFTR is similar to several cation channel proteins (*27*) and some cation-translocating adenosine triphosphatases (ATPases) (*28*)

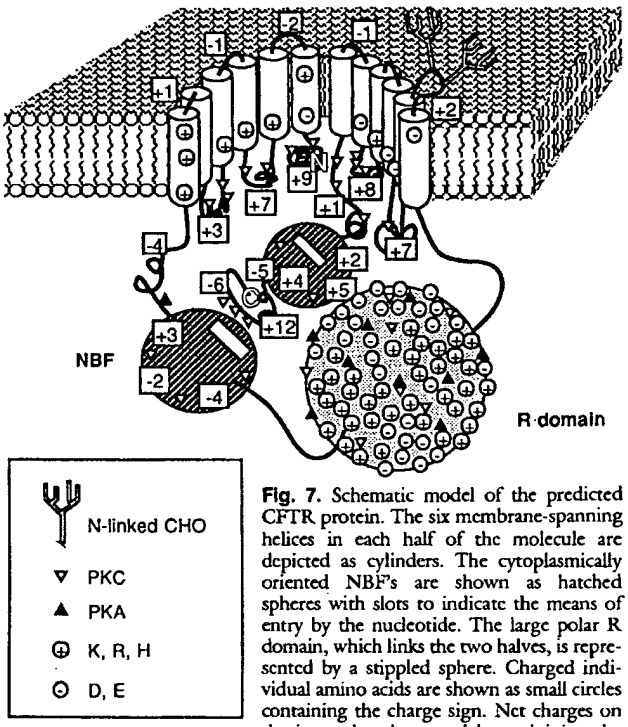


Fig. 7. Schematic model of the predicted CFTR protein. The six membrane-spanning helices in each half of the molecule are depicted as cylinders. The cytoplasmically oriented NBF's are shown as hatched spheres with slots to indicate the means of entry by the nucleotide. The large polar R domain, which links the two halves, is represented by a stippled sphere. Charged individual amino acids are shown as small circles containing the charge sign. Net charges on the internal and external loops joining the membrane cylinders and on regions of the NBF's are contained in open squares. Potential sites for phosphorylation by protein kinases A or C (PKA or PKC) and N-glycosylation (N-linked CHO) are as indicated. K, Lys; R, Arg; H, His; D, Asp; and E, Glu.

as well as the recently described adenylate cyclase of bovine brain (*29*). Short regions of sequence identity have also been detected between the putative transmembrane regions of CFTR and other membrane-spanning proteins (*30*). In addition, a sequence of 18 amino acids situated approximately 50 residues from the carboxyl terminus of CFTR shows some identity (12/18) with the raf serine-threonine kinase proto-oncogene product of *Xenopus laevis* (*31*).

Finally, a sequence identity (10 of 13 amino acid residues) has been noted between a hydrophilic segment (position 701 to 713) within the highly charged R domain of CFTR and a region immediately preceding the first transmembrane loop of the sodium channels in both rat brain and eel (*32*). This feature of CFTR is not shared with the topologically closely related P-glycoprotein; the 241–amino acid linking peptide is apparently the major difference between the two proteins.

Relevance to the CF anion transport defect. In view of the genetic data of Kerem *et al.* (*9*) and the tissue specificity and predicted properties of the CFTR protein, it is reasonable to conclude that CFTR is directly responsible for CF. It remains unclear, however, how CFTR is involved in the regulation of ion conductance across the apical membrane of epithelial cells.

It is possible that CFTR serves as an ion channel itself. For example, 10 of the 12 putative transmembrane regions contain one or more amino acids with charged side chains (Fig. 7), a property similar to that of the brain sodium channel and the γ-aminobutyric acid (GABA) receptor chloride channel subunits, where charged residues are present in four of the six, and three of the four, respective membrane-associated domains per subunit or repeat unit (*32, 33*). The amphipathic nature of these transmembrane segments is believed to contribute to the channel-forming capacity of these molecules. In contrast, the closely related P-glycoprotein, which is

not believed to conduct ions, has only two charged residues in all 12 transmembrane domains. Alternatively, CFTR may not be an ion channel but instead it may serve to regulate ion channel activities. In support of the latter possibility, none of the recently purified polypeptides (from trachea and kidney) that are capable of reconstituting chloride channels in lipid membranes (*6*) appear to be CFTR, judged on the basis of molecular mass.

In any case, the presence of ATP-binding domains in CFTR suggests that ATP hydrolysis is directly involved and required for the transport function. The high density of phosphorylation sites for protein kinases A and C and the clusters of charged residues in the R domain may both serve to regulate this activity. The deletion of Phe^{508} in the NBF may prevent proper binding of ATP or the conformational change required for normal CFTR activity, consequently resulting in the observed insensitivity to activation by protein kinase A– or protein kinase C–mediated phosphorylation of the CF apical chloride conductance pathway (*5*). Since the predicted structure of CFTR contains several conserved domains and belongs to a family of proteins, most of which function as parts of multicomponent molecular systems (*15*), the CFTR protein may also participate in epithelial cell functions not related to ion transport.

To understand the basic defect in CF, it is necessary to determine the precise role of Phe^{508} in the regulation of ion transport and to understand the mechanism that leads to the pathophysiology of the disease. With the CF gene (that is, the cDNA) now isolated, it should be possible to elucidate the control of ion transport pathways in epithelial cells in general. Knowledge gained from study of the CF gene product (CFTR), both the normal and mutant forms, will provide a molecular basis for the development of improved means of treatment of the disease.

REFERENCES AND NOTES

1. T. F. Boat, M. J. Welsh, A. L. Beaudet, in *The Metabolic Basis of Inherited Disease*, C. L. Scriver, A. L. Beaudet, W. S. Sly, D. Valle, Eds. (McGraw-Hill, New York, ed. 6, 1989), pp. 2649–2680.
2. M. R. Knowles *et al.*, *Science* **221**, 1067 (1983); M. R. Knowles, J. Gatzy, R. Boucher, *J. Clin. Invest.* **71**, 1410 (1983); P. M. Quinton, *Nature* **301**, 421 (1983); K. Sato, *Am. J. Physiol.* **247**, R646 (1984).
3. J. H. Widdicombe, M. J. Welsh, W. E. Finkbeiner, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 6167 (1985); P. M. Quinton and J. Bijman, *N. Engl. J. Med.* **308**, 1185 (1983).
4. J. R. Yankaskas, C. U. Cotton, M. R. Knowles, J. T. Gatzy, R. C. Boucher, *Am. Rev. Resp. Dis.* **132**, 1281 (1985); P. S. Pedersen, E. H. Larsen, B. Hainau, N. J. Brandt, *Med. Sci. Res.* **15**, 1009 (1987); T. J. Jensen and J. R. Riordan, unpublished results.
5. M. J. Welsh, *Science* **232**, 1648 (1986); R. A. Frizzell, G. Rechkemmer, R. L. Shoemaker, *ibid.* **233**, 558 (1986); M. J. Welsh and C. M. Liedtke, *Nature* **322**, 467 (1986); R. A. Schoumacher *et al.*, *ibid.* **330**, 752 (1987); M. Li *et al.*, *ibid.* **331**, 358 (1988); T.-C. Hwang *et al.*, *Science* **244**, 1351 (1989); M. Li *et al.*, *ibid.*, p. 1353.
6. D. W. Landry *et al.*, *Science* **244**, 1469 (1989).
7. J. M. Rommens *et al.*, *ibid.* **245**, 1059 (1989).
8. M. Farrall *et al.*, *Am. J. Hum. Genet.* **43**, 471 (1988).
9. B. Kerem *et al.*, *Science* **245**, 1073 (1989).
10. The cDNA libraries from cultured epithelial cells were prepared as follows: sweat gland cells derived from a non-CF individual and from a CF patient were grown to first passage as described [G. Collie, M. Buchwald, P. Harper, J. R. Riordan, *In Vitro Cell. Dev. Biol.* **21**, 592 (1985)]. The presence in these cells of an outwardly rectifying Cl^- channel was confirmed (J. A. Tabcharani, T. J. Jensen, J. R. Riordan, J. W. Hanrahan, *J. Membrane Biol.*, in press), but the CF cells were insensitive to activation by cyclic AMP [T. J. Jensen, J. W. Hanrahan, J. A. Tabcharani, M. Buchwald, J. R. Riordan, *Pediatric Pulmonol. Suppl.* **2**, 100 (1988)]. Polyadenylated RNA was isolated [J. M. Chirgwin, A. E. Przybyla, R. J. Macdonald, W. J. Rutter, *Biochemistry* **18**, 5294 (1979); H. Aviv and P. Leder, *Proc. Natl. Acad. Sci. U.S.A.* **69**, 1408 (1972)] and used as template for the synthesis of cDNA according to U. Gubler and B. Hoffman [*Gene* **25**, 263 (1983)]. After methylation of internal Eco RI sites, ends were made flush with T4 DNA polymerase, and phosphorylated Eco RI linkers were added to the cDNA. After digestion with Eco RI and removal of excess linkers, the cDNA products were ligated into the Eco RI site of λ ZAP (Stratagene, San Diego, CA). The same procedures were used to construct a library from RNA isolated from preconfluent cultures of the T84 colon carcinoma cell line [K. Dharmsathaphorn, J. A. McRoberts, K. G. Mandel, L. D. Tisdale, H. Masui, *Am. J. Physiol.* **246**, G204 (1984)]. The numbers of independent recombinants in the three libraries were: 2.0×10^6 for the non-CF sweat gland cells, 4.5×10^6 for the CF sweat gland cells, and 3.2×10^6 from T84 cells. Standard procedures were used for screening [T. Maniatis, E. F. Fritsch, J. Sambrook, *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1982)]. Bluescript plasmids were rescued from plaque-purified clones by excision with M13 helper phage (Stratagene). The lung and pancreas libraries were purchased from Clontech Lab Inc. (Catalog Nos. HL1066b and HL1069h, respectively).
11. The start point of the CF gene transcript was derived by primer extension procedures [F. J. Calzone, R. J. Britten, E. H. Davidson, *Methods Enzymol.* **152**, 611 (1987)]. The oligonucleotide primer [positioned 157 nucleotides (nt) from the 5′ end of the 10-1 clone] was end-labeled with [γ-^{32}P]ATP (Amersham, 5000 Ci/mmole) and T4 polynucleotide kinase, purified by gel filtration, and annealed with ~5 μg of T84 poly(A)$^+$ RNA for 2 hours at 60°C. The extension reaction was performed at 41°C for 1 hour with avian myeloblastosis virus (AMV) reverse transcriptase (Life Sciences, Inc.) and terminated by addition of NaOH to 0.4*M* and EDTA to 20 m*M*, with subsequent neutralization with ammonium acetate (*p*H 4.6). The products were treated with phenol, precipitated with ethanol, redissolved in buffer with formamide, and analyzed on a polyacrylamide sequencing gel.
12. The anchored PCR procedure [M. A. Frohman, M. K. Dush, G. R. Martin, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 8998 (1988)] was used to synthesize cDNA corresponding to the 5′ and 3′ ends of the transcript. For the 5′ end clones, poly(A)$^+$ RNA from pancreas and T84 cells were subjected to reverse-transcription with the use of an exon 2–specific primer (*11*). The first strand cDNA products were fractionated on an agarose column and the fractions containing large species were identified by gel electrophoresis after the polymerase chain reaction [R. K. Saiki *et al.*, *Science* **230**, 1350 (1985)] with a pair of oligonucleotide primers (145 nt apart within the 10-1 sequence) just 5′ of the extension primer. These products were pooled, concentrated, and treated with terminal deoxynucleotidyl transferase (BRL) and dATP, as recommended by the supplier. Second strand synthesis was performed with Taq Polymerase (Cetus, AmpliTaq) and an oligonucleotide containing a linker sequence, 5′-CGGAATTCTCGAGATC(T)$_{12}$-3′. This linker, together with another primer (internal to the extension primer) with an Eco RI restriction site at its 5′ end, was then used for PCR. After digestion with Eco RI and Bgl II, products were purified and cloned in Bluescript KS (Stratagene) by standard procedures. All the recovered clones contained inserts of more than 350 nt. The 3′ end clones were generated with the use of similar procedures. PCR amplification was carried out with the linker described above and an oligonucleotide with the sequence 5′-ATGAAGTCCAAGGATTTAG-3′, which is ~70 nt upstream of the Hind III site at position 5027 (Fig. 2). The products were digested with Hind III and Xho I and cloned in the Bluescript vector. Candidate clones were identified by hybridization with the 3′ end of cDNA T16-4.5. All PCR's were performed for 30 cycles as described by the enzyme supplier.
13. M. Kozak, *Nucleic Acids Res.* **12**, 857 (1984); *Cell* **44**, 283 (1986).
14. Other sequence differences were noted between the normal (T16-4.5) and CF (C1-1/5) cDNA clones. At position 2629 (Fig. 2), T16-4.5 contained a C and C1-1/5 a T, resulting in a change of Leu to Phe. At position 4555, the base was G in T16-4.5 but A in C1-1/5 (Val to Met). The differences may be results of cDNA cloning artifacts or may represent sequence polymorphisms. Specific oligonucleotide hybridization analysis of patient or family DNA should distinguish these possibilities. Since these changes are conserved amino acid substitutions, they are unlikely to be causative mutations. Additional nucleotide differences were observed in the 3′ untranslated region between different cDNA clones and the corresponding genomic DNA sequence.
15. C. F. Higgins, M. P. Gallagher, M. L. Mimmack, S. R. Pearce, *BioEssays* **8**, 111 (1988); C. F. Higgins *et al.*, *Nature* **323**, 448 (1986).
16. P. Gros, J. Croop, D. E. Housman, *Cell* **47**, 371 (1986); C. Chen *et al.*, *ibid.*, p. 381; J. H. Gerlach *et al.*, *Nature* **324**, 485 (1986); P. Gros, M. Raymond, J. Bell, D. Housman, *Mol. Cell. Biol.* **8**, 2770 (1988).
17. Several large families of integral membrane proteins are known, including: (i) A number of ligand-gated ion channels of which the nicotinic acetylcholine receptor [R. M. Stroud and J. Finer-Moore, *Annu. Rev. Cell Biol.* **1**, 317 (1985)] is the prototype. Receptors for the inhibitory neurotransmitters GABA (*33*) and glycine are included in this family. (ii) A family of ion channels with a totally different structural motif are the voltage-gated, sodium, calcium, and potassium channels (*27*). (iii) Involved in the translocation of ions are the structurally related cation pumps such as the Ca^{2+}-ATPase [C. J. Brandl, N. M. Green, B. Korczak, D. H. MacLennan, *Cell* **44**, 597 (1986)], the Na^+,K^+-ATPase [G. E. Shull and J. B. Lingrel, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 4039 (1987)], and the H^+,K^+-ATPase [G. E. Shull and J. B. Lingrel, *J. Biol. Chem.* **261**, 16788 (1986)]. These are but examples.
18. G. B. Wilson, T. L. Jahn, J. R. Fonseca, *Clin. Chim. Acta* **49**, 79 (1973); V. van Heyningen, C. Hayward, J. Fletcher, C. McAuley, *Nature* **315**, 513 (1985).
19. J. Garnier, D. J. Osguthorpe, B. Robson, *J. Mol. Biol.* **120**, 97 (1978); J. Kyte and R. F. Doolittle, *ibid.* **157**, 105 (1982).
20. D. D. Pless and W. J. Lennarz, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 134 (1977).
21. P. J. Blackshear, A. C. Nairn, J. F. Kuo, *FASEB J.* **2**, 2957 (1989).
22. J. Hess, W. Wels, M. Vogel, W. Goebel, *FEMS Microbiol. Lett.* **34**, 1 (1986).
23. K. O'Hare, C. Murphy, R. Levis, G. M. Rubin, *J. Mol. Biol.* **180**, 437 (1984).
24. M. J. Friedrich, L. C. DeVeaux, R. J. Kadner, *J. Bacteriol.* **167**, 928 (1986).
25. K. Ohyama *et al.*, *Nature* **322**, 572 (1986).
26. G. F.-L. Ames, *Annu. Rev. Biochem.* **55**, 397 (1986).
27. M. Noda *et al.*, *Nature* **320**, 188 (1986); T. Tanabe *et al.*, *ibid.* **328**, 313 (1987); A. Baumann *et al.*, *EMBO J.* **6**, 3419 (1987).
28. C.-M. Chen, T. K. Misra, S. Silver, B. P. Rosen, *J. Biol. Chem.* **261**, 15030 (1986).
29. J. Krupinski *et al.*, *Science* **244**, 1558 (1989).
30. In addition to the major NBF homologies, searches of the PIR and SWISSPROT data bases detected shorter stretches of sequence homology with other proteins including the following: